



King's Research Portal

Document Version
Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Walsh, S. L. F., Wells, A. U., Desai, S. R., Poletti, V., Piciocchi, S., Dubini, A., Nunes, H., Valeyre, D., Brillet, P., Kambouchner, M., Morais, A., Pereira, J. M., Souto Moura, C. P., Grutters, J. C., van den Heuvel, D. A., van Es, H. W., van Oosterhout, M. F., Seldenrijk, C. A., Bendstrup, E., ... Hansell, D. (2016). Multicentre evaluation of multidisciplinary team meeting agreement on diagnosis in diffuse parenchymal lung disease: a case-cohort study. *The Lancet Respiratory Medicine*.

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Manuscript Number: THELANCETRM-D-16-00035R3

Title: Multicentre evaluation of multidisciplinary team meeting agreement on diagnosis in diffuse parenchymal lung disease: a case cohort study

Article Type: Articles (Original Research)

Keywords: multidisciplinary team, diagnosis, diffuse parenchymal lung disease, idiopathic pulmonary fibrosis, interobserver agreement

Corresponding Author: Dr. Simon LF Walsh, MD MRCP FFRCSI

Corresponding Author's Institution: KING'S COLLEGE

First Author: Simon LF Walsh, MD MRCP FFRCSI

Order of Authors: Simon LF Walsh, MD MRCP FFRCSI; Athol U Wells, MD FRCP FRCP; Sujal R Desai, MD FRCP FRCP; Venerino Poletti, MD PhD; Sara Piciucchi; Alessandra Dubini; Hilario Nunes, MD, PhD; Dominique Valeyre, MD; Pierre Y Brillet, MD, PhD; Marianne Kambouchner, MD; António Morais, MD, PhD; José M Pereira, MD; Conceição P Souto Moura, MD; Jan C Grutters, MD, PhD, Prof; Daniel A van den Heuvel, MD; Hendrik W van Es, MD, PhD; Matthijs F van Oosterhout, MD, PhD; Cornelis A Seldenrijk, MD, PhD; Elisabeth Bendstrup, MD, PhD; Finn Rasmussen, MD; MDSc; Line B Madsen, MD, PhD; Bibek Goopu, PhD, MRCP; Sabine Pomplun, FRCPath; Hiroyuki Taniguchi, MD, PhD; Junya Fukuoka, MD, PhD.; Takeshi Johkoh, M.D., Ph.D.; Andrew G Nicholson, DM FRCPath; Charlie Sayer, MBBS FRCP; Lilian Edmunds, FRCPath; Joseph Jacob, MBBS(Hos), MRCP, MRCS, DTM&H, FRCP; Maria A Kokosi, MD; Jeffrey L Myers, MD; Kevin R Flaherty, MD, MS; David M Hansell, MD FRCP FRCP FRSM

Manuscript Region of Origin: UNITED KINGDOM

Abstract: Background

To evaluate inter-multidisciplinary team agreement for the diagnosis of diffuse parenchymal lung disease (DPLD).

Methods

Seven multidisciplinary meetings (MDTMs) consisting of at least one clinician, radiologist and pathologist, from 7 different countries evaluated 70 cases of diffuse lung disease in a two-stage process. First, the clinician, radiologist and pathologist (when lung biopsy was performed) evaluated each case and chose likelihoods (censored at 5% and summing to 100% in each case) for each of their differential diagnoses, without inter-disciplinary consultation. A full MDTM with review of all clinical, radiologic and pathologic data followed this. Interobserver agreement and inter-MDTM agreement for diagnosis was calculated using Cohen's kappa coefficient or weighted kappa coefficient where appropriate.

Findings

Inter-MDTM agreement for first choice diagnoses was acceptable ($\kappa = 0.50$). Idiopathic pulmonary fibrosis made up 18% of all MDTM first choice diagnoses. Diagnostic likelihoods for MDTM differential diagnoses were converted to a 5-point scale (0 = condition not included in the differential diagnosis, 1 = low probability (5-25%), 2 = intermediate probability (30-65%), 3 = high probability (70-95%), and 4 = pathognomonic (100%)). Inter-MDTM agreement on diagnostic likelihoods was good for idiopathic pulmonary fibrosis (IPF) ($\kappa_w = 0.71$) and connective tissue disease related interstitial lung disease (CTD-ILD) ($\kappa_w = 0.73$), only moderate for non-specific interstitial pneumonia (NSIP) ($\kappa_w = 0.42$) and poor for hypersensitivity pneumonitis (HP) ($\kappa_w = 0.29$). MDTMs, clinicians and radiologists respectively gave high confidence diagnoses of IPF (>65% likelihood) in 77.3%, 64.6% and 66.3% of cases. The prognostic significance of a first choice diagnosis of IPF versus not IPF was evaluated for MDTMs, clinicians and radiologists. Greater prognostic significance was demonstrated for an MDTM diagnosis of IPF as compared to individual clinician's diagnosis of IPF in 5/7 MDTMs, radiologist's diagnosis of IPF in 4/7 MDTMs.

Interpretation

Agreement between MDTMs for diagnosis in diffuse lung disease is acceptable and good for a diagnosis of IPF. This is validated by the greater prognostic significance of an IPF diagnosis made by MDTMs as compared to individual clinicians or radiologists. Furthermore, MDTMs made the diagnosis of IPF with higher confidence and more frequently than clinicians or radiologists. MDTM agreement for diagnosis of NSIP and hypersensitivity pneumonitis is poor, indicating a need for international consensus on diagnostic criteria for these diseases.

**Multicentre evaluation of multidisciplinary team meeting agreement on
diagnosis in diffuse parenchymal lung disease: a case cohort study**

Simon LF Walsh MD

Prof Athol U Wells MD

Sujal R Desai MD

Prof Venerino Poletti, MD

Sara Piciucchi MD

Alessandra Dubini MD

Prof Hilario Nunes, MD

Prof Dominique Valeyre, MD

Prof Pierre Y Brillet, MD

Marianne Kambouchner, MD

Prof António Morais, MD

José M Pereira, MD

Conceição P Souto Moura, MD

Prof Jan C Grutters, MD

Daniel A van den Heuvel, MD

Hendrik W van Es, MD

Matthijs F van Oosterhout, MD

Cornelis A Seldenrijk, MD

Elisabeth Bendstrup PhD

Finn Rasmussen, MD

Line B Madsen, PhD

Bibek Gooptu, PhD

Sabine Pomplun, FRCPath

Hiroyuki Taniguchi, MD

Prof Junya Fukuoka, MD

Takeshi Johkoh, MD

Prof Andrew G Nicholson, DM

Charlie Sayer FRCR

Lilian Edmunds, FRCPath

Joseph Jacob FRCR

Maria A Kokosi, MD

Prof Jeffrey L Myers, MD

Prof Kevin R Flaherty MD

Prof David M Hansell FRSM

Abstract

Background

To evaluate inter-multidisciplinary team agreement for the diagnosis of diffuse parenchymal lung disease (DPLD).

Methods

Seven multidisciplinary meetings (MDTMs) consisting of at least one clinician, radiologist and pathologist, from 7 different countries evaluated 70 cases of diffuse lung disease in a two-stage process. First, the clinician, radiologist and pathologist (when lung biopsy was performed) evaluated each case and chose likelihoods (censored at 5% and summing to 100% in each case) for each of their differential diagnoses, without inter-disciplinary consultation. A full MDTM with review of all clinical, radiologic and pathologic data followed this. Interobserver agreement and inter-MDTM agreement for diagnosis was calculated using Cohen's kappa coefficient or weighted kappa coefficient where appropriate.

Findings

Inter-MDTM agreement for first choice diagnoses was acceptable ($\kappa = 0.50$).

Idiopathic pulmonary fibrosis made up 18% of all MDTM first choice diagnoses.

Diagnostic likelihoods for MDTM differential diagnoses were converted to a 5-point scale (0 = condition not included in the differential diagnosis, 1 = low probability (5–25%), 2 = intermediate probability (30–65%), 3 = high probability (70–95%), and 4 = pathognomonic (100%)). Inter-MDTM agreement on diagnostic likelihoods was good for idiopathic pulmonary fibrosis (IPF) ($\kappa_w = 0.71$) and connective tissue disease related interstitial lung disease (CTD-ILD) ($\kappa_w = 0.73$), only moderate for non-specific

interstitial pneumonia (NSIP) (κ = 0.42) and poor for hypersensitivity pneumonitis (HP) (κ = 0.29). MDTMs, clinicians and radiologists respectively gave high confidence diagnoses of IPF (>65% likelihood) in 77.3%, 64.6% and 66.3% of cases. The prognostic significance of a first choice diagnosis of IPF versus not IPF was evaluated for MDTMs, clinicians and radiologists. Greater prognostic separation was demonstrated for an MDTM diagnosis of IPF as compared to individual clinician's diagnosis of IPF in 5/7 MDTMs, radiologist's diagnosis of IPF in 4/7 MDTMs.

Interpretation

Agreement between MDTMs for diagnosis in diffuse lung disease is acceptable and good for a diagnosis of IPF. This is validated by the greater prognostic separation of an IPF diagnosis made by MDTMs as compared to individual clinicians or radiologists. Furthermore, MDTMs made the diagnosis of IPF with higher confidence and more frequently than clinicians or radiologists. MDTM agreement for diagnosis of NSIP and hypersensitivity pneumonitis is poor, indicating a need for international consensus on diagnostic criteria for these diseases.

Funding

National Institute of Health Research Respiratory Disease Biomedical Unit at the Royal Brompton and Harefield NHS Foundation Trust and Imperial College London DMH is the recipient of a National Institute of Health Research Senior Investigator Award.

Introduction

Diffuse parenchymal lung disease (DPLD) represents a diverse and challenging group of pulmonary disorders with varied prognoses and different management options. A consistent diagnostic approach to these diseases is essential if clinical trial data is to be reliably applied to individual patients. With the recent licensing of two new anti-fibrotic idiopathic pulmonary fibrosis (IPF) therapies, accurate and consistent diagnosis of IPF is of particular importance if clinical benefit is to be achieved^{1,2}. In 2002, a joint statement by the American Thoracic Society (ATS) and the European Respiratory Society (ERS) on the classification of the idiopathic interstitial pneumonias (IIPs) advocated a multidisciplinary diagnostic approach involving integration of clinical, radiologic and, when lung biopsy material is available, pathologic data³. This approach has been emphasised by several studies in recent years and was re-stated in the 2013 ATS/ERS update on the IIPs⁴⁻⁸. Although this recommendation specifically applies to IIP, the multidisciplinary approach has been widely adopted as the diagnostic gold standard for DPLD in general^{5,7}. Several studies have evaluated interobserver agreement for diagnosis in the setting of DPLD^{5-7,9}, but most predate the 2013 ATS/ERS IIP classification update, the 2011 ATS/ERS/JRS/ALAT statement on the diagnosis and management of IPF and the availability of novel anti-fibrotic IPF drugs, all of which may impact decisions on diagnosis^{1,2,4,10}. Furthermore, many of these studies focus on individual observers rather than agreement between multidisciplinary teams (MDTs)^{5-7,9,11}. In this study, we evaluated the level of inter-MDTM diagnostic agreement between seven international centres.

Methods

Patient and Multidisciplinary team selection

For the retrospective examination of clinically indicated data, the institutional ethics review board waived informed patient consent. Patients selected for this study represented consecutive patients presenting to the interstitial lung disease unit of the Royal Brompton & Harefield NHS Foundation Trust and requiring MDTM characterisation, between March 1st 2010 and August 31st 2010. Only patients who had all their clinical investigations (serology, HRCT and when required, surgical lung biopsy) performed at the host institution were included. A total of 7 MDTMs from 7 different countries (Denmark, France, Italy, Japan, Netherlands, Portugal and the United Kingdom), each with specialist expertise in the diagnosis and management of DPLD, were invited to participate in the study. The only prerequisite for participation was that each MDT had a regular multidisciplinary meeting for DPLD in place with consistent attendance by at least one clinician, radiologist and pathologist.

Evaluation of cases

The evaluation of each of case took place in two stages.

1. First, clinicians, radiologists and pathologists were required to review the cases independently without inter-specialty consultation. Clinicians had access to all the presenting clinical information (age, gender, smoking history, history of established connective tissue disease (CTD), symptoms including symptoms suggestive of a CTD, autoantibody profile, exposure history, medications at presentation, bronchoalveolar lavage result if performed, and ACE level if performed), pulmonary function tests and HRCT (without access to the original

HRCT report). Radiologists and pathologists had access only to the age, gender and smoking history for the patient, and the HRCT (radiologist) or digitized surgical lung biopsy slides (pathologist) at presentation. Specifically, pathologists had access to all the pathology data that was available in the form of digitized slides (in .svs format) and were viewed using Aperio ImageScope 12.3 viewing software. This digital viewing application has all the imaging functionality normally available to pathologists in routine clinical practice and is used by the host institution to evaluate cases referred from outside institutions for opinions.

For each patient, observers were required to select up to 5 differential diagnoses and provide a diagnostic likelihood (censored at 5% and summing to 100% in each case) from a drop-down menu of diffuse lung diseases (Appendix, Table A1). The only stipulation was that diagnoses were considered in the context of the current ATS/ERS classification and terminology for the IIPs⁴.

2. Second, the clinician, radiologist and pathologist convened as an MDT and reviewed the cases together, again providing up to 5 diagnoses with diagnostic likelihoods (also censored at 5% and summing to 100% in each case). All the clinical information supplied in the first stage, pulmonary function tests and HRCT at presentation as well as digitized surgical lung biopsy slides were available to the MDT for review.

Outcome

As a means of validating the diagnosis made by MDTM versus individual specialists, the mortality of each groups' diagnosis of IPF was compared. This was achieved by separating the entire cohort into a binary IPF diagnosis category (IPF and not IPF) for each MDTM, clinician and radiologist based upon assigned diagnoses. The survival period for each patient was calculated from the date of referral to the host institution to the minimum of date of death, date the patient was last known to be alive or 1st June 2015 (end of the study period). Vital status for each patient on the 1st June 2015 was obtained by evaluating his or her electronic patient record.

Statistical analysis

Data are given as means with standard deviations (SD), medians with interquartile range (IQR) or as number of patients and percentage where appropriate. Statistical analyses were performed using STATA (version 12, StataCorp, College Station, Texas).

Cohen's kappa coefficient (κ) was used to evaluate interobserver and inter-MDTM agreement for diagnosis. Cohen's weighted kappa coefficient (κ_w) was used to evaluate interobserver agreement and inter-MDTM agreement for an estimation of the probability of each diagnosis. In order to do this, the percentage diagnostic likelihood given for each diagnosis was converted to a 5 point scale (0–4), representing clinically useful probabilities: 0 = condition not included in the differential diagnosis, 1 = low probability (5–25%), 2 = intermediate probability (30–65%), 3 = high probability (70–95%), and 4 = pathognomonic (100%). For example, if the differential diagnoses given by an MDT were IPF (65% diagnostic likelihood), NSIP

(25% diagnostic likelihood) and hypersensitivity pneumonitis (10% diagnostic likelihood), the probability grades for IPF, NSIP and hypersensitivity pneumonitis for this case would be 2, 1 and 1 respectively. Weighted kappa values were calculated between paired observers (for statements of interobserver agreement), and between paired MDTs (for statements of inter-MDTM agreement) and expressed as median values with interquartile ranges for all unique combinations of pairs (21 for 7 observers or 7 MDTs). Weighting the Kappa coefficient allowed the degree of disagreement to be quantified by assigning greater emphasis to large differences between scores. Weighted kappa coefficients were categorized as follows: poor ($0 < \kappa_w \leq 0.20$), fair ($0.20 < \kappa_w \leq 0.40$), moderate ($0.40 < \kappa_w \leq 0.60$), good ($0.60 < \kappa_w \leq 0.80$) and excellent ($0.80 < \kappa_w \leq 1.00$). This approach has been used in previous investigations of interobserver agreement for diagnosis in diffuse lung diseases^{9,11}.

In addition to the above, in each case the first choice diagnosis was considered “low confidence” (diagnostic likelihood <70%), or “high confidence” (70% or greater diagnostic likelihood). These categories were based on the diagnostic likelihood categories used to assess the clinical probability of pulmonary embolism in the PIOPED study¹².

Univariate Cox regression analysis was used to identify associations between mortality and MDTM, clinician and radiologist diagnoses in terms of IPF versus “not IPF”. Reported hazards ratios are for diagnosis of IPF versus “not IPF”. The assumptions of proportional hazards were tested by 1) visual inspection the log-log plot of survival, 2) comparison of the Kaplan-Meier observed survival curves with the Cox predicted curves for the same variable and 3) graphical and formal analysis of Schoenfeld residuals (analysis not shown). Results are reported as hazard ratios, p-values and 95% confidence intervals and graphically as Kaplan-Meier survival curves.

Role of the funding source:

The sponsors of the study did not have any role in the design, data collection, analysis and interpretation, nor in the writing of the report. The corresponding author had full access to all of the data and the final responsibility to submit for publication.

Results

Patient population

A total of 113 consecutive new referrals who required local MDTM characterisation were identified from the clinical database between March 1st 2010 and August 31st 2010. Of these, 43/113 (38.1%) were excluded on the basis that a) their initial work-up HRCT scan had been performed at the referring institution (n=29), b) their lung function had been performed at the referring institution (n=4), or c) surgical lung biopsy had been performed by the referring institution (n=10), (Table 1 and Appendix, Figure A1). The remaining 70 cases made up the final study population. Basic patient demographics are shown in the Appendix, Table A2. Of note, thirteen (19%) had an established diagnosis of a connective tissue disease at the time of presentation to the host institution and 22/70 (34.1%) underwent surgical lung biopsy at the host institution. In cases where surgical lung biopsy was not performed (48/70), a confident diagnosis had been made without the need for surgical lung biopsy material. Vital status was known for all patients at the end of the study period. These 70 patients resulted in the assignment of 490 first choice MDTM diagnoses (70 patients evaluated by 7 MDTMs). First choice diagnoses are shown in Table 2. The NSIP/OP overlap ILD category was combined with the NSIP category and diagnosis categories

whose frequency was less than 10% of the total number of first choice diagnoses were combined into an 'others' diagnosis category. The final diagnosis categories were then as follows: CTD related ILD (n=146/490, 29.8%), IPF (n=88/490, 18.0%), idiopathic NSIP (n=50/490, 10.2%), hypersensitivity pneumonitis (n=46/490, 9.4%) and others (n=160/490, 32.7%). Subsequent analyses focused on these four primary diagnosis categories.

Inter-multidisciplinary team and inter-observer agreement for first choice diagnosis

Inter-MDTM agreement and interobserver agreement (for clinicians, radiologists and pathologists) for first choice diagnosis is shown in Table 3. Overall inter-MDTM agreement for first choice diagnosis was moderate ($\kappa = 0.50$). Inter-MDTM agreement for a first choice diagnosis of IPF was good ($\kappa = 0.60$), for a first choice diagnosis of CTD related ILD also good ($\kappa = 0.64$), but poor for first choice diagnoses of idiopathic NSIP ($\kappa = 0.25$) and hypersensitivity pneumonitis ($\kappa = 0.24$). On subgroup analysis in patients in whom lung biopsy was not performed (n=48/70, 68.6%), overall inter-MDTM agreement for diagnosis was greater ($\kappa = 0.57$) with inter-MDTM agreement for first choice diagnoses of IPF, CTD related ILD and hypersensitivity pneumonitis also greater. Overall inter-observer agreement among clinicians for first choice diagnosis was moderate ($\kappa = 0.45$) and fair between radiologists ($\kappa = 0.33$) and pathologists ($\kappa = 0.31$) (Table 3.).

Inter-multidisciplinary team and inter-observer agreement for diagnosis probabilities

There was fair to good inter-MDTM agreement on the estimation of diagnostic likelihood of the 4 most prevalent diagnoses as shown in Table 4. In particular, inter-MDTM agreement on the diagnostic likelihood of IPF was good (κ = 0.71, IQR 0.64-0.77), and good for CTD related ILD (κ = 0.73, IQR 0.68-0.78) but moderate for idiopathic NSIP (κ = 0.41 IQR 0.37-0.49) and fair for hypersensitivity pneumonitis (κ = 0.29 IQR 0.24-0.40). Subgroup analysis of inter-MDTM agreement on the estimation of diagnostic likelihood of IPF in patients without lung biopsy was good (κ = 0.78 IQR 0.74-0.83).

Agreement between clinicians on the probability of a diagnosis of IPF or CTD-related ILD was superior to agreement on the probability of a diagnosis of idiopathic NSIP or hypersensitivity pneumonitis. Agreement between radiologists or pathologists on the probability of a diagnosis of IPF was superior to agreement on the probability of a diagnosis of CTD-related ILD, idiopathic NSIP or hypersensitivity pneumonitis.

Subgroup analysis of inter-MDTM agreement in patients without an established diagnosis of a connective tissue disease.

At the time of patient selection, 13/70 (18.6%) patients had an established diagnosis of a CTD (systemic sclerosis = 7, rheumatoid arthritis = 3, Sjögren's syndrome = 2, mixed connective tissue disease = 1). In order to investigate if high agreement in these 13 cases caused a spurious increase in agreement on non-CTD diagnoses and in particular, impacting agreement on a diagnosis of IPF, a subgroup analysis was performed in the remainder of the cohort (n=57/70, 81.4%). On this analysis, although inter-MDTM agreement for a first choice diagnosis of CTD related ILD decreased (κ = 0.42), no significant change in inter-MDTM agreement was observed for a first choice

diagnosis of IPF ($\kappa = 0.58$), idiopathic NSIP ($\kappa = 0.24$) or hypersensitivity pneumonitis ($\kappa = 0.23$) (Appendix, Table A2).

Diagnostic confidence for first choice diagnoses

A total of 347/490 (70.1%) first choice MDT diagnoses were made with high confidence (diagnostic likelihood 70–95% = high confidence, 100% = pathognomonic). The median prevalence of first choice MDT diagnoses made with high confidence was 67.1% (IQR 54.3–88.8). The median prevalence of first choice diagnoses made with high confidence by clinicians, radiologists or pathologists was 58.9% (IQR 52.9–71.4), 68.6% (IQR 35.7–85.7) and 72.7% (IQR 59.1–81.8) respectively (Appendix, Table A3). On subgroup analysis in the 48/70 (68.6%) patients who did not undergo surgical lung biopsy 237/336 (70.5%) first choice MDT diagnoses were made with high confidence. Within this subgroup the median prevalence of first choice diagnoses made with high confidence by the MDTs, clinicians or radiologists were 68.7% (IQR 52.8–87.5), 60.4% (IQR 37.5–75.0) and 66.6% (IQR 39.6–83.3) respectively.

For the diagnosis of IPF, supportive non-significant trends for higher confidence diagnoses by MDTs (68/88, 77.3%) as compared to clinicians (62/96, 64.6%) or radiologists (57/86, 66.3%) were demonstrated ($p = 0.23$). In the 22/70 (31.4%) cases that underwent surgical lung biopsy (therefore a total of 154 diagnoses rendered by 7 pathologists), 15/154 (9.7%) cases were assigned a diagnosis of IPF, of which 12/15 (80.0%) were assigned with high confidence.

A review of the cases where the pathologists gave a first choice diagnosis of IPF (15/154 cases, 154 = 22 cases x 7 pathologists) was performed to ascertain if, in cases where surgical lung biopsy was performed and the first choice

pathologic diagnosis was IPF, the final MDTM diagnosis was usually IPF. In 6/15 cases, despite the pathologist giving a first choice diagnosis of IPF, the final MDTM first choice diagnosis was not IPF. Furthermore, in only 2/15 cases, was IPF not already suggested by either the clinician or radiologist in that MDTM (Table A5).

Evaluation of the prognostic significance of an MDTM diagnosis of IPF

On univariate Cox regression analysis, the multidisciplinary distinction between IPF and other diagnoses demonstrated non-significant trends toward greater prognostic separation (as judged by hazard ratio p values) than the clinician distinction (in 5/7 groups) or the radiologist distinction (in 4/7 groups) (Table 5). A graphical presentation of the KM-survival curves for the categorisation of first-choice diagnosis as IPF or not IPF in at least 4/7 MDTMs, 4/7 clinicians and 4/7 radiologists is shown in Figure 1. The same analysis for pathologists' diagnosis of IPF failed to reach statistical significance for 5/7 pathologists, probably because of the small subgroup size (n=22) and low prevalence of IPF within this subgroup (15/154, pathologists' first-choice diagnoses were IPF) (Appendix, Table A4 and A5).

Discussion

We have demonstrated for the first time, that there is acceptable diagnostic agreement between multidisciplinary groups in the setting of diffuse parenchymal lung disease and that this agreement is validated by the greater prognostic significance of an IPF diagnosis made by multidisciplinary groups as compared to individual clinicians and radiologists. Furthermore, MDTMs make the diagnosis of IPF with high confidence, more frequently than clinicians or radiologists.

Since the publication of the ATS/ERS 2002 consensus statement on the classification of the IIPs, multidisciplinary evaluation of DPLD has been widely adopted as the diagnostic gold standard^{3,4}. This diagnostic approach has been investigated in a limited way in several settings. Flaherty et al. examined the formulation of diagnosis in a cohort of diffuse lung diseases against interobserver agreement and diagnostic confidence within a single multidisciplinary team (consisting of 3 clinicians, 2 radiologists and 2 pathologists) and demonstrated that diagnostic agreement between observers improved with successive integration of clinical, radiologic and pathologic data⁵. In a second study, Flaherty et al. expanded these findings by demonstrating higher levels of agreement between academic physicians, radiologists and pathologists for diagnosis in diffuse lung disease when compared with their community counterparts⁷. Several years later, Thomeer et al. demonstrated in a cohort of patients included in an IPF trial, high accuracy for a clinical diagnosis of IPF made by 6 respiratory physicians from different European centres⁶. A limitation of these studies is that not all evaluated agreement between different MDTs for diagnosis⁵, one focuses specifically on the diagnosis of IPF⁶ and all of them predate the most recent 2013 ATS/ERS IIP update^{4,7,9}. The current study

is the first to evaluate inter-MDTM agreement for diagnosis in diffuse lung disease since publication of the 2013 ATS/ERS update on the classification of the IIPs.

With the recent licensing of two new drugs for treatment of IPF in mild to moderate disease, early and accurate diagnosis of IPF is crucial. Our study demonstrates that inter-MDTM agreement for the diagnosis of IPF is good, with clinicians demonstrating only marginally inferior levels of agreement, as compared to MDTMs for this diagnosis. On subgroup analysis in patients without surgical lung biopsy, inter-MDTM agreement and interobserver agreement between clinicians for the diagnosis of IPF improved and were again, almost equal. As evidence-based diagnostic criteria for IPF are now clearly defined and in many IPF patients, are relatively easily applied, this near parity of agreement between MDTMs and between clinicians is not surprising^{4,10}. For example, in a patient with classic appearances for usual interstitial pneumonia (UIP) on HRCT, a rapidly progressive disease course and no identifiable triggers, multidisciplinary discussion is unlikely to alter a clinician's initial impression of IPF. Our findings suggest that formal multidisciplinary input may not be necessary in every case of suspected IPF if expert clinical evaluation is available, which may be of particular relevance to centres with limited access to appropriate radiology or pathology expertise¹³.

A strength of our findings is that the greater agreement on multidisciplinary diagnosis is mirrored by the trend toward greater prognostic separation of a multidisciplinary distinction between IPF and other diagnoses when compared to the clinicians' or radiologists' distinction. To demonstrate this, patients were necessarily selected from 2010 to allow an analysis of 5-year survival. An added advantage of this approach was that these patients were referred in a pre-antifibrotic drug era, therefore mortality differences between IPF and non-IPF patients were not

confounded by antifibrotic therapy. The fact that trends are present but inconclusive may reflect powering limitations but may also indicate that multidisciplinary discussion adds prognostic value in the sub-group of patients in which there is significant diagnostic uncertainty

In contrast to those with IPF, many patients with non-IPF interstitial lung diseases are not covered by evidence-based diagnostic guidelines therefore diagnosis is driven by clinical reasoning and analysis of all available data in a multidisciplinary setting. In these situations, a level of disagreement between MDTMs is predictable and borne out by the poor level of inter-MDTM agreement in the current study for diagnoses of NSIP and hypersensitivity pneumonitis^{4,14-16}. In particular, it is well recognised that the diagnosis of hypersensitivity pneumonitis can be challenging as it relies on an array of data, none of which is definitive when considered in isolation¹⁵, and at least on HRCT, patterns of NSIP, UIP or organising pneumonia may be the sole expressions of this disease¹⁶⁻¹⁹. The poor MDTM agreement for the diagnosis of hypersensitivity pneumonitis in the current study highlights an urgent need for international consensus on what hypersensitivity pneumonitis actually is.

Our use of the weighted kappa to investigate inter-MDTM and interobserver agreement on diagnostic probabilities is similar to other studies of diagnostic agreement, but warrants particular consideration⁹. Converting the diagnostic likelihoods to a 5-point probability scale enabled examination of the range of diagnostic likelihoods from minimal likelihood to pathognomonic using the weighted kappa. As excluding IPF is as important as making a diagnosis of IPF, this methodology has allowed us to demonstrate that there is good agreement on the likelihood of IPF and as stated previously, this reflects consistent application of established diagnostic guidelines for this disease. In contrast, MDTM agreement on

the likelihood of hypersensitivity pneumonitis was poor, reinforcing the view that MDTMs were unclear on how this diagnosis is made.

Two separate observations from our study warrant further discussion. First, 13 patients had an established diagnosis of a connective tissue disease at presentation. However, following MDT evaluation, a CTD diagnosis was constructed in an additional 8 cases or more by five of the seven MDTMs based on presenting clinical symptoms and serology. Separating patients with IIP from those with CTD related ILD can be challenging - some patients present with subtle clinical features or serological abnormalities which suggest an autoimmune process but do not meet established criteria for a specific CTD ²⁰⁻²². Recently an ERS/ATS task force was formed in order to establish consensus on how to classify these patients and a set of diagnostic criteria have been suggested ²³⁻²⁵. Following removal of patients with an established diagnosis of CTD, our subgroup analysis demonstrated acceptable levels of agreement on CTD related ILD (in contrast to hypersensitivity pneumonitis), underlying the importance of reaching international consensus on definitions. Furthermore, the high frequency of CTD related ILD diagnoses made in the current study highlights the importance of formal rheumatology input within the MDTM, which might necessarily include face-to-face rheumatologic consultation with patients, suspected of having an undiagnosed connective tissue disease.

Second, it has previously been suggested that dynamic exchanges of clinical, radiologic and pathologic information between experts in a multidisciplinary process results in higher confidence diagnoses⁵. In the current study however, the proportion of high confidence diagnoses (70% or greater) assigned following MDTM evaluation did not increase when compared to the proportion of high confidence diagnoses being assigned by the individual components of the MDTM. In the majority of MDTMs, high

confidence diagnoses were more frequently assigned by radiologists and pathologists, when compared to MDTM diagnoses or diagnoses assigned by clinicians. As radiologists and pathologists in the current study did not have access to clinical information, their interpretation was based almost entirely on pattern recognition, which might conceivably result in more confident, but not necessarily more accurate diagnoses. Although one benefit of the multidisciplinary process is that diagnoses may be challenged and must be publically defended, it is possible that extra discussion creates more difficulty in some cases that initially seem straightforward when evaluated by individuals in isolation. It must be highlighted however, that in the specific case of IPF, MDTMs made the diagnosis of IPF with high confidence more frequently than clinicians or radiologists.

Our methodology has some limitations. As patients were selected from a pre-antifibrotic drug era, IPF was possibly not as prevalent in our study group as it would be now at most referral centres which expend more time in evaluating patients for approval for anti-fibrotic treatment. However as discussed, this allowed us to evaluate the veracity of MDTM diagnosis for IPF, against outcome. Second, unlike a real-world multidisciplinary process, none of the observers had face-to-face consultation with patients and therefore did not have the benefit of obtaining a clinical history or performing physical examination first-hand. In difficult to characterize patients, a clinician's impression might conceivably be influenced by direct contact with the patient in ways, which are not easily and objectively quantified. Equally important, but also difficult to assess, is the impact of individual personalities on MDTM diagnosis. Arguably a consensus diagnosis at multidisciplinary evaluation might sometimes reflect the strongest voice in the room rather than represent true agreement between participants. Lastly, the wide range of proportions of high confidence diagnoses

demonstrated between the different MDTM's seen in our study may reflect cultural influences on the dynamics of multidisciplinary evaluation. These less tangible influences could be the focus of further studies.

In conclusion, our study demonstrates that diagnostic agreement between MDTMs is superior to interobserver agreement between clinicians, radiologists and pathologists in the setting of DPLD. In particular, inter-MDTM agreement for a diagnosis of IPF is good and validated by the greater prognostic significance of an IPF diagnosis made by multidisciplinary groups as compared to individual specialists. Poor levels of inter-MDTM agreement were demonstrated for NSIP and hypersensitivity pneumonitis, the latter of which may relate to the lack of evidence-based guidelines for diagnosing this disease.

Research in context

Evidence before this study

We performed a PubMed search on 17th January 2014 using the search terms “diffuse parenchymal lung disease”, “idiopathic pulmonary fibrosis”, “idiopathic interstitial pneumonias”, “interobserver agreement”, “diagnosis” and “multidisciplinary team” for the period between January 2000 and January 2014. This search was extended to 1st December 2015 during the writing of the manuscript. Our search was restricted to publications written in English. We identified 7 key publications which were pertinent to our study^{3-7,9,10}. Of these, 4 studies of observer agreement in setting of diffuse parenchymal lung disease were identified^{5-7,9}. All of these studies however predated the most up to date ATS/ERS update on the classification of the idiopathic interstitial pneumonias⁴, and not all of them evaluated diagnostic

agreement between multidisciplinary teams but rather focused on diagnostic agreement between individual observers.

Added value of this study

This study is the first evaluation of inter-multidisciplinary team agreement for diagnosis in the setting of diffuse parenchymal lung disease since the updated 2013 ATS/ERS classification of the idiopathic interstitial pneumonias and the 2011 ATS/ERS/JRS/ALAT guidelines for the diagnosis and management of idiopathic pulmonary fibrosis.

Implications of all the available evidence

Our study demonstrates that diagnostic agreement between MDTMs is superior to interobserver agreement between clinicians, radiologists and pathologists in the setting of DPLD. Specifically in IPF, MDTMs have a higher level of agreement on diagnosis, assign diagnosis with higher confidence more frequently, and provide diagnoses that have greater prognostic significance than clinicians or radiologists in the majority of groups.

This is of particular importance, as accurate and consistent diagnosis of idiopathic pulmonary fibrosis is needed if clinical outcome is to be optimised. In contrast, inter-multidisciplinary team agreement for a diagnosis of hypersensitivity pneumonitis is poor, highlighting an urgent need for standardised diagnostic guidelines for this disease entity.

Conflicts of interest:

AUW received personal fees from Intermune, Roche, Bayer and Gilead. DV received personal fees from Roche, Intermune and Boehringer Ingelheim. HN is an investigator for clinical trials by Intermune, Roche, Boehringer-Ingelheim, Sanofi, and Centocor. HT received personal fees from Abbott Japan Co. Ltd, Actelion Pharmaceuticals Japan Ltd, Ashai Kasei Pharmaceutical Corporation, Astellas Pharmaceutical Incorporated, Astra Zeneca, Bayer, Boehringer Ingelheim, and Chugai Pharmaceuticals. JF received personal fees from Astellas Pharmaceutical Incorporated, Pathology Institute Corporation, Chugai Pharmaceuticals and Sakura Finetek Japan. AGN received personal fees from Sanofi, Intermune, Boehringer Ingelheim and Actelion. KRF received personal fees from Boehringer Ingelheim, Genentech, Ikaria, Immuneworks, Veracyte, Roche, Gilead, Biogen, Afferent, Aeolus, Pharmakea and grants from Boehringer Ingelheim, Genentech, Roche and Afferent. DMH received personal fees from Boehringer Ingelheim, Sanofi, AstraZeneca, Roche and Glaxo Smith Kline. The other authors declared no conflicts of interest.

Author contributions

LFW, AUW, DMH, SD - Study concept, data collection, analysis and manuscript writing. JJ, MK, CS - Data collection. AGN, LE – Data collection and evaluation of digitized pathology data. KRF, JLM – Manuscript review and editing. Scoring of MDT data. All remaining authors made up the 7 multidisciplinary teams from France, Italy, Netherlands, Japan, Portugal, United Kingdom, Denmark

1. King TE, Jr., Bradford WZ, Castro-Bernardini S, et al. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 2014; **370**(22): 2083-92.
2. Richeldi L, du Bois RM, Raghu G, et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med* 2014; **370**(22): 2071-82.
3. American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. This joint statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was adopted by the ATS board of directors, June 2001 and by the ERS Executive Committee, June 2001. *Am J Respir Crit Care Med* 2002; **165**(2): 277-304.
4. Travis WD, Costabel U, Hansell DM, et al. An official American Thoracic Society/European Respiratory Society statement: Update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. *Am J Respir Crit Care Med* 2013; **188**(6): 733-48.
5. Flaherty KR, King TE, Jr., Raghu G, et al. Idiopathic interstitial pneumonia: what is the effect of a multidisciplinary approach to diagnosis? *Am J Respir Crit Care Med* 2004; **170**(8): 904-10.
6. Thomeer M, Demedts M, Behr J, et al. Multidisciplinary interobserver agreement in the diagnosis of idiopathic pulmonary fibrosis. *Eur Respir J* 2008; **31**(3): 585-91.
7. Flaherty KR, Andrei AC, King TE, Jr., et al. Idiopathic interstitial pneumonia: do community and academic physicians agree on diagnosis? *Am J Respir Crit Care Med* 2007; **175**(10): 1054-60.

8. Tomassetti S, Piciocchi S, Tantalocco P, Dubini A, Poletti V. The multidisciplinary approach in the diagnosis of idiopathic pulmonary fibrosis: a patient case-based review. *Eur Respir Rev* 2015; **24**(135): 69-77.
9. Aziz ZA, Wells AU, Hansell DM, et al. HRCT diagnosis of diffuse parenchymal lung disease: inter-observer variation. *Thorax* 2004; **59**(6): 506-11.
10. Raghu G, Collard HR, Egan JJ, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011; **183**(6): 788-824.
11. Walsh SL, Calandriello L, Sverzellati N, Wells AU, Hansell DM, Consort UIPO. Interobserver agreement for the ATS/ERS/JRS/ALAT criteria for a UIP pattern on CT. *Thorax* 2015.
12. Investigators P. Value of the ventilation/perfusion scan in acute pulmonary embolism. Results of the prospective investigation of pulmonary embolism diagnosis (PIOPED). *JAMA* 1990; **263**(20): 2753-9.
13. Jo HE, Corte TJ, Moodley Y, et al. Evaluating the interstitial lung disease multidisciplinary meeting: a survey of expert centres. *BMC Pulm Med* 2016; **16**(1): 22.
14. Travis WD, Hunninghake G, King TE, Jr., et al. Idiopathic nonspecific interstitial pneumonia: report of an American Thoracic Society project. *Am J Respir Crit Care Med* 2008; **177**(12): 1338-47.
15. Lacasse Y, Selman M, Costabel U, et al. Clinical diagnosis of hypersensitivity pneumonitis. *Am J Respir Crit Care Med* 2003; **168**(8): 952-8.
16. Silva CI, Muller NL, Lynch DA, et al. Chronic hypersensitivity pneumonitis: differentiation from idiopathic pulmonary fibrosis and nonspecific interstitial pneumonia by using thin-section CT. *Radiology* 2008; **246**(1): 288-97.

17. Lacasse Y, Girard M, Cormier Y. Recent advances in hypersensitivity pneumonitis. *Chest* 2012; **142**(1): 208-17.
18. Churg A, Sin DD, Everett D, Brown K, Cool C. Pathologic patterns and survival in chronic hypersensitivity pneumonitis. *Am J Surg Pathol* 2009; **33**(12): 1765-70.
19. Silva CI, Churg A, Muller NL. Hypersensitivity pneumonitis: spectrum of high-resolution CT and pathologic findings. *AJR Am J Roentgenol* 2007; **188**(2): 334-44.
20. Fischer A, du Bois R. Interstitial lung disease in connective tissue disorders. *Lancet* 2012; **380**(9842): 689-98.
21. Tzelepis GE, Toya SP, Moutsopoulos HM. Occult connective tissue diseases mimicking idiopathic interstitial pneumonias. *Eur Respir J* 2008; **31**(1): 11-20.
22. Mittoo S, Gelber AC, Christopher-Stine L, Horton MR, Lechtzin N, Danoff SK. Ascertainment of collagen vascular disease in patients presenting with interstitial lung disease. *Respir Med* 2009; **103**(8): 1152-8.
23. Fischer A, West SG, Swigris JJ, Brown KK, du Bois RM. Connective tissue disease-associated interstitial lung disease: a call for clarification. *Chest* 2010; **138**(2): 251-6.
24. Vij R, Noth I, Strek ME. Autoimmune-featured interstitial lung disease: a distinct entity. *Chest* 2011; **140**(5): 1292-9.
25. Kinder BW, Collard HR, Koth L, et al. Idiopathic nonspecific interstitial pneumonia: lung manifestation of undifferentiated connective tissue disease? *Am J Respir Crit Care Med* 2007; **176**(7): 691-7.

Variable	Value
Age (Years)	60.9 ± 15.5
Gender (M : F)	24:46
Smoking (Never/ever/current)	38:24:8
Established CTD history	13
Biopsied (Y/N)	22/48
DL_{co} (% Predicted)	44.8 ± 14.5
FEV₁ (% Predicted)	73.0 ± 20.5
FVC (% Predicted)	79.0 ± 19.6

Table 1. Patient demographics. CTD = connective tissue disease. Where appropriate, values are expressed as means ± SD.

Diagnosis	Number (n=490)
Connective tissue disease related ILD	146 (29.8%)
IPF	88 (18.0%)
Hypersensitivity pneumonitis	46 (9.6%)
NSIP	42 (9.4%)
Unclassifiable ILD	38 (7.8%)
Other	37 (7.6%)
Sarcoidosis	20 (4.1%)
Drug-related ILD	18 (3.7%)
Smoking-related ILD	16 (3.3%)
Occupational lung disease	14 (2.9%)
Pleuroparenchymal fibroelastosis	10 (2.0%)
OP/NSIP overlapping disease	8 (1.6%)
Organizing pneumonia	7 (1.4%)

Table 2. First choice diagnoses given by 7 multidisciplinary teams for 70 cases of diffuse lung disease. ILD = interstitial lung disease, NSIP =non-specific interstitial pneumonia, OP = organising pneumonia.

	MDTM (κ)	Clinicians (κ)	Radiologists (κ)	Pathologists (κ)
Overall Total (No biopsy)	0.50 (0.57)	0.45 (0.50)	0.33 (0.31)	0.31
IPF Total (No biopsy)	0.60 (0.70)	0.59 (0.71)	0.46 (0.42)	0.46
NSIP Total (No biopsy)	0.25 (0.25)	0.19 (0.19)	0.25 (0.25)	0.23
CTD Total (No biopsy)	0.64 (0.73)	0.57 (0.62)	0.10 (0.11)	0.22
HP Total (No biopsy)	0.24 (0.31)	0.25 (0.38)	0.27 (0.22)	0.20

Table 3. Unweighted Kappa values (κ) for inter-MDTM agreement for overall diagnoses and for individual diagnoses of idiopathic pulmonary fibrosis (IPF), connective tissue disease related interstitial lung disease (CTD), non-specific interstitial pneumonia (NSIP) and hypersensitivity pneumonitis (HP). Unweighted Kappa values for inter-observer agreement among clinicians, radiologists, and pathologists are also shown. Figures in parentheses are Kappa values in non-biopsied cases, n=48 (see text). *22/70 (31.4%) cases had surgical lung biopsy performed.

	MDTM (κw)	Clinicians (κw)	Radiologists (κw)	Pathologists (κw)
IPF	0.71 (0.64-0.77)	0.72 (0.67-0.76)	0.60 (0.46-0.66)	0.58 (0.45-0.66)
CTD	0.73 (0.68-0.78)	0.76 (0.70-0.78)	0.17 (0.08-0.31)	0.21 (0.06-0.36)
NSIP	0.42 (0.37-0.49)	0.31 (0.27-0.41)	0.32 (0.26-0.41)	0.30 (0.00-0.53)
HP	0.29 (0.24-0.40)	0.42 (0.30-0.47)	0.35 (0.29-0.43)	0.26 (0.10-0.45)

Table 4. Weighted Kappa values (**κw**) for estimation of diagnostic likelihood for individual diagnoses of idiopathic pulmonary fibrosis (IPF), connective tissue disease related interstitial lung disease (CTD), non-specific interstitial pneumonia (NSIP) and hypersensitivity pneumonitis (HP), expressed as medians with interquartile ranges.

	MDTM (HR, p-value, 95%CI)	Clinicians (HR, p-value, 95%CI)	Radiologists (HR, p-value, 95%CI)
Team 1	2.67, p = 0.016, 1.21-6.02	2.09, p = 0.085, 0.90-4.86	2.80, p = 0.021, 1.17-6.73
Team 2	3.44, p = 0.003, 1.54-7.68	2.95, p = 0.008, 1.33-6.59	4.08, p = 0.001, 1.84-9.04
Team 3	5.30, p < 0.001, 2.26-12.41	3.75, p = 0.002, 1.65-8.51	2.78, p = 0.030, 1.11-6.97
Team 4	3.99, p = 0.006, 1.49-10.66	3.34, p = 0.007, 1.38-8.00	4.49, p = 0.003, 1.71-12.29
Team 5	2.61, p = 0.025, 1.12-6.06	2.03, p = 0.100, 0.87-4.69	2.58, p = 0.033, 1.08-6.21
Team 6	3.36, p = 0.007, 1.40-8.07	4.14, p = 0.002, 1.72-9.97	2.11, p = 0.082, 0.91-4.89
Team 7	2.43, p = 0.030, 1.09-5.41	2.96, p = 0.007, 1.43-6.55	1.28, p = 0.583, 0.53-3.06

Table 5. Univariate Cox regression analysis for mortality according to MDTM, clinician and radiologists diagnoses of IPF versus not IPF. (MDTM = multidisciplinary team meeting). The results for the MDTMs, clinicians and radiologists are based upon the full patient cohort (n=70). *The results for the pathologists are for the subgroup of 22/70 (31.4%) patients who underwent surgical lung biopsy

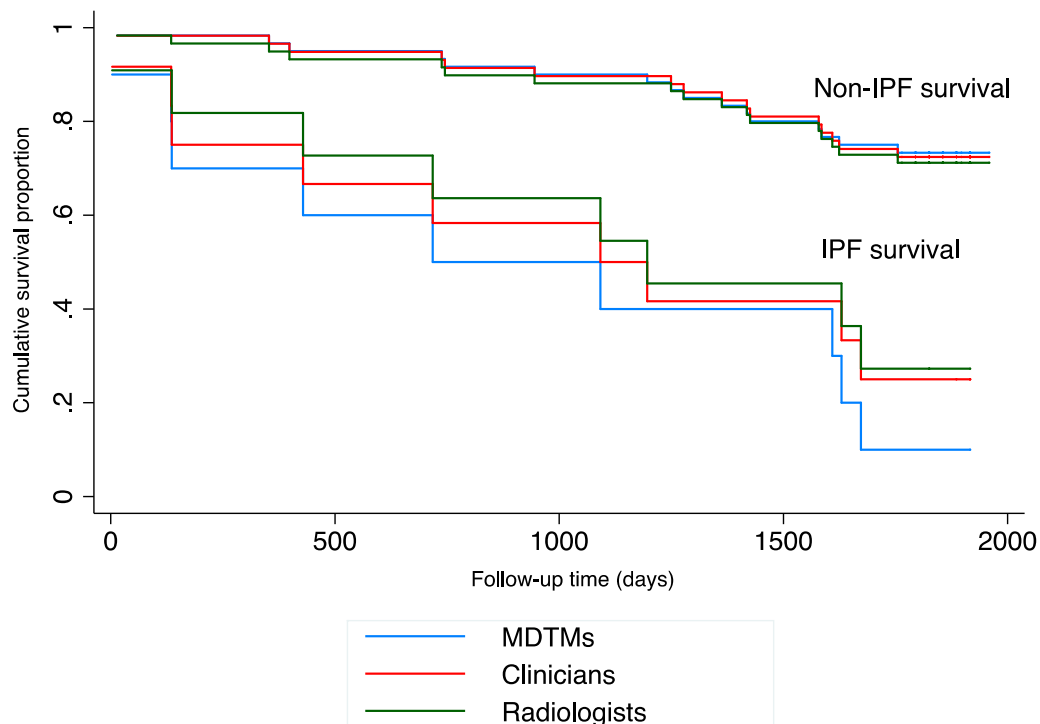


Figure 1. Kaplan-Meier survival curve demonstrating survival differences between patients assigned a diagnosis of IPF or not IPF. The entire cohort was separated into cases where at least 4/7 MDTMs assigned a first-choice diagnosis of IPF (Blue KM curve, HR 6.26, $p < 0.0001$, 95% CI 2.72-14.33), at least 4/7 clinicians assigned a first-choice diagnosis of IPF (red KM curve, HR 4.43, $p < 0.0001$, 95% CI 1.94-10.01), and at least 4/7 radiologists assigned a first-choice diagnosis of IPF (green KM curve, HR 3.76, $p = 0.002$, 95% CI 1.61-8.75). MDTM = multidisciplinary team meeting. See Appendix, Figure 7A for number at risk table.

**MA-multicentre evaluation of multidisciplinary team meeting agreement on
diagnosis in diffuse parenchymal lung disease: a case cohort study**

Simon LF Walsh MD

Prof Athol U Wells MD

Sujal R Desai MD

Prof Venerino Poletti, MD

Sara Piciucchi MD

Alessandra Dubini MD

Prof Hilario Nunes, MD

Prof Dominique Valeyre, MD

Prof Pierre Y Brillet, MD

Marianne Kambouchner, MD

Prof António Morais, MD

José M Pereira, MD

Conceição P Souto Moura, MD

Prof Jan C Grutters, MD

Daniel A van den Heuvel, MD

Hendrik W van Es, MD

Matthijs F van Oosterhout, MD

Cornelis A Seldenrijk, MD

Elisabeth Bendstrup PhD

Finn Rasmussen, MD

Line B Madsen, PhD

Bibek Gooptu, PhD

Sabine Pomplun, FRCPath

Hiroyuki Taniguchi, MD

Prof Junya Fukuoka, MD

Takeshi Johkoh, MD

Prof Andrew G Nicholson, DM

Charlie Sayer FRCR

Lilian Edmunds, FRCPath

Joseph Jacob FRCR

Maria A Kokosi, MD

Prof Jeffrey L Myers, MD

Prof Kevin R Flaherty MD

Prof David M Hansell FRSM

Abstract

Background

To evaluate inter-multidisciplinary team agreement for the diagnosis of diffuse parenchymal lung disease (DPLD).

Methods

Seven multidisciplinary meetings (MDTMs) consisting of at least one clinician, radiologist and pathologist, from 7 different countries evaluated 70 cases of diffuse lung disease in a two-stage process. First, the clinician, radiologist and pathologist (when lung biopsy was performed) evaluated each case and chose likelihoods (censored at 5% and summing to 100% in each case) for each of their differential diagnoses, without inter-disciplinary consultation. A full MDTM with review of all clinical, radiologic and pathologic data followed this. Interobserver agreement and inter-MDTM agreement for diagnosis was calculated using Cohen's kappa coefficient or weighted kappa coefficient where appropriate.

Findings

Inter-MDTM agreement for first choice diagnoses was acceptable ($\kappa = 0.50$). Idiopathic pulmonary fibrosis made up 18% of all MDTM first choice diagnoses. Diagnostic likelihoods for MDTM differential diagnoses were converted to a 5-point scale (0 = condition not included in the differential diagnosis, 1 = low probability (5–25%), 2 = intermediate probability (30–65%), 3 = high probability (70–95%), and 4 = pathognomonic (100%)). Inter-MDTM agreement on diagnostic likelihoods was good for idiopathic pulmonary fibrosis (IPF) ($\kappa_w = 0.71$) and connective tissue disease related interstitial lung disease (CTD-ILD) ($\kappa_w = 0.73$), only moderate for non-specific

interstitial pneumonia (NSIP) (κ = 0.42) and poor for hypersensitivity pneumonitis (HP) (κ = 0.29). MDTMs, clinicians and radiologists respectively gave high confidence diagnoses of IPF (>65% likelihood) in 77.3%, 64.6% and 66.3% of cases. The prognostic significance of a first choice diagnosis of IPF versus not IPF was evaluated for MDTMs, clinicians and radiologists. Greater prognostic [separationsignificance](#) was demonstrated for an MDTM diagnosis of IPF as compared to individual clinician's diagnosis of IPF in 5/7 MDTMs, radiologist's diagnosis of IPF in 4/7 MDTMs.

Interpretation

Agreement between MDTMs for diagnosis in diffuse lung disease is acceptable and good for a diagnosis of IPF. This is validated by the greater prognostic [significance separation](#) of an IPF diagnosis made by MDTMs as compared to individual clinicians or radiologists. Furthermore, MDTMs made the diagnosis of IPF with higher confidence and more frequently than clinicians or radiologists. MDTM agreement for diagnosis of NSIP and hypersensitivity pneumonitis is poor, indicating a need for international consensus on diagnostic criteria for these diseases.

Funding

National Institute of Health Research Respiratory Disease Biomedical Unit at the Royal Brompton and Harefield NHS Foundation Trust and Imperial College London DMH is the recipient of a National Institute of Health Research Senior Investigator Award.

Introduction

Diffuse parenchymal lung disease (DPLD) represents a diverse and challenging group of pulmonary disorders with varied prognoses and different management options. A consistent diagnostic approach to these diseases is essential if clinical trial data is to be reliably applied to individual patients. With the recent licensing of two new anti-fibrotic idiopathic pulmonary fibrosis (IPF) therapies, accurate and consistent diagnosis of IPF is of particular importance if clinical benefit is to be achieved^{1,2}. In 2002, a joint statement by the American Thoracic Society (ATS) and the European Respiratory Society (ERS) on the classification of the idiopathic interstitial pneumonias (IIPs) advocated a multidisciplinary diagnostic approach involving integration of clinical, radiologic and, when lung biopsy material is available, pathologic data³. This approach has been emphasised by several studies in recent years and was re-stated in the 2013 ATS/ERS update on the IIPs⁴⁻⁸. Although this recommendation specifically applies to IIP, the multidisciplinary approach has been widely adopted as the diagnostic gold standard for DPLD in general^{5,7}. Several studies have evaluated interobserver agreement for diagnosis in the setting of DPLD^{5-7,9}, but most predate the 2013 ATS/ERS IIP classification update, the 2011 ATS/ERS/JRS/ALAT statement on the diagnosis and management of IPF and the availability of novel anti-fibrotic IPF drugs, all of which may impact decisions on diagnosis^{1,2,4,10}. Furthermore, many of these studies focus on individual observers rather than agreement between multidisciplinary teams (MDTs)^{5-7,9,11}. In this study, we evaluated the level of inter-MDTM diagnostic agreement between seven international centres.

Methods

Patient and Multidisciplinary team selection

For the retrospective examination of clinically indicated data, the institutional ethics review board waived informed patient consent. Patients selected for this study represented consecutive patients presenting to the interstitial lung disease unit of the Royal Brompton & Harefield NHS Foundation Trust and requiring MDTM characterisation, between March 1st 2010 and August 31st 2010. Only patients who had all their clinical investigations (serology, HRCT and when required, surgical lung biopsy) performed at the host institution were included. A total of 7 MDTMs from 7 different countries (Denmark, France, Italy, Japan, Netherlands, Portugal and the United Kingdom), each with specialist expertise in the diagnosis and management of DPLD, were invited to participate in the study. The only prerequisite for participation was that each MDT had a regular multidisciplinary meeting for DPLD in place with consistent attendance by at least one clinician, radiologist and pathologist.

Evaluation of cases

The evaluation of each of case took place in two stages.

1. First, clinicians, radiologists and pathologists were required to review the cases independently without inter-specialty consultation. Clinicians had access to all the presenting clinical information (age, gender, smoking history, history of established connective tissue disease (CTD), symptoms including symptoms suggestive of a CTD, autoantibody profile, exposure history, medications at presentation, bronchoalveolar lavage result if performed, and ACE level if

performed), pulmonary function tests and HRCT (without access to the original HRCT report). Radiologists and pathologists had access only to the age, gender and smoking history for the patient, and the HRCT (radiologist) or digitized surgical lung biopsy slides (pathologist) at presentation. Specifically, pathologists had access to all the pathology data that was available in the form of digitized slides (in .svs format) and were viewed using Aperio ImageScope 12.3 viewing software. This digital viewing application has all the imaging functionality normally available to pathologists in routine clinical practice and is used by the host institution to evaluate cases referred from outside institutions for opinions.

For each patient, observers were required to select up to 5 differential diagnoses and provide a diagnostic likelihood (censored at 5% and summing to 100% in each case) from a drop-down menu of diffuse lung diseases (Appendix, Table A1). The only stipulation was that diagnoses were considered in the context of the current ATS/ERS classification and terminology for the IIPs⁴.

2. Second, the clinician, radiologist and pathologist convened as an MDT and reviewed the cases together, again providing up to 5 diagnoses with diagnostic likelihoods (also censored at 5% and summing to 100% in each case). All the clinical information supplied in the first stage, pulmonary function tests and HRCT at presentation as well as digitized surgical lung biopsy slides were available to the MDT for review.

Outcome

As a means of validating the diagnosis made by MDTM versus individual specialists, the mortality of each groups' diagnosis of IPF was compared. This was achieved by separating the entire cohort into a binary IPF diagnosis category (IPF and not IPF) for each MDTM, clinician and radiologist based upon assigned diagnoses. The survival period for each patient was calculated from the date of referral to the host institution to the minimum of date of death, date the patient was last known to be alive or 1st June 2015 (end of the study period). Vital status for each patient on the 1st June 2015 was obtained by evaluating his or her electronic patient record.

Statistical analysis

Data are given as means with standard deviations (SD), medians with interquartile range (IQR) or as number of patients and percentage where appropriate. Statistical analyses were performed using STATA (version 12, StataCorp, College Station, Texas).

Cohen's kappa coefficient (κ) was used to evaluate interobserver and inter-MDTM agreement for diagnosis. Cohen's weighted kappa coefficient (κ_w) was used to evaluate interobserver agreement and inter-MDTM agreement for an estimation of the probability of each diagnosis. In order to do this, the percentage diagnostic likelihood given for each diagnosis was converted to a 5 point scale (0–4), representing clinically useful probabilities: 0 = condition not included in the differential diagnosis, 1 = low probability (5–25%), 2 = intermediate probability (30–65%), 3 = high probability (70–95%), and 4 = pathognomonic (100%). For example, if the

differential diagnoses given by an MDT were IPF (65% diagnostic likelihood), NSIP (25% diagnostic likelihood) and hypersensitivity pneumonitis (10% diagnostic likelihood), the probability grades for IPF, NSIP and hypersensitivity pneumonitis for this case would be 2, 1 and 1 respectively. Weighted kappa values were calculated between paired observers (for statements of interobserver agreement), and between paired MDTs (for statements of inter-MDTM agreement) and expressed as median values with interquartile ranges for all unique combinations of pairs (21 for 7 observers or 7 MDTs). Weighting the Kappa coefficient allowed the degree of disagreement to be quantified by assigning greater emphasis to large differences between scores. Weighted kappa coefficients were categorized as follows: poor ($0 < \kappa_w \leq 0.20$), fair ($0.20 < \kappa_w \leq 0.40$), moderate ($0.40 < \kappa_w \leq 0.60$), good ($0.60 < \kappa_w \leq 0.80$) and excellent ($0.80 < \kappa_w \leq 1.00$). This approach has been used in previous investigations of interobserver agreement for diagnosis in diffuse lung diseases^{9,11}.

In addition to the above, in each case the first choice diagnosis was considered “low confidence” (diagnostic likelihood <70%), or “high confidence” (70% or greater diagnostic likelihood). These categories were based on [these the diagnostic likelihood categories](#) used to assess the clinical probability of pulmonary embolism in the PIOPED study¹².

Univariate Cox regression analysis was used to identify associations between mortality and MDTM, clinician and radiologist diagnoses in terms of IPF versus “not IPF”. [Reported hazards ratios are for diagnosis of IPF versus “not IPF”](#). The assumptions of proportional hazards were tested by 1) visual inspection the log-log plot of survival, 2) comparison of the Kaplan-Meier observed survival curves with the Cox predicted curves for the same variable and 3) graphical and formal analysis of Schoenfeld residuals ([analysis not shown](#)). Results are reported as hazard ratios, p-

Formatted: Indent: First line: 1.27 cm

values and 95% confidence intervals and graphically as Kaplan-Meier survival curves.

Role of the funding source:

The sponsors of the study did not have any role in the design, data collection, analysis and interpretation, nor in the writing of the report. The corresponding author had full access to all of the data and the final responsibility to submit for publication.

Results

Patient population

A total of 113 consecutive new referrals who required local MDTM characterisation were identified from the clinical database between March 1st 2010 and August 31st 2010. Of these, 43/113 (38.1%) were excluded on the basis that a) their initial work-up HRCT scan had been performed at the referring institution (n=29), b) their lung function had been performed at the referring institution (n=4), or c) surgical lung biopsy had been performed by the referring institution (n=10), (Table 1 and Appendix, Figure A1). The remaining 70 cases made up the final study population. Basic patient demographics are shown in the Appendix, Table A2. Of note, thirteen (19%) had an established diagnosis of a connective tissue disease at the time of presentation to the host institution and 22/70 (34.1%) underwent surgical lung biopsy at the host institution. In cases where surgical lung biopsy was not performed (48/70), a

Formatted: Font: (Default) Arial, 12 pt

Formatted: Level 1, Line spacing: Double

Formatted: Font: (Default) Arial, 12 pt

Formatted: Line spacing: Double

Formatted: Font: (Default) Arial, 12 pt

Formatted: Font: (Default) Arial, 12 pt

Formatted: Font: (Default) Arial, 12 pt

Formatted: Indent: First line: 0 cm

[confident diagnosis had been made without the need for surgical lung biopsy material.](#)

Vital status was known for all patients at the end of the study period. These 70 patients resulted in the assignment of 490 first choice MDTM diagnoses (70 patients evaluated by 7 MDTMs). First choice diagnoses are shown in Table 2. The NSIP/OP overlap ILD category was combined with the NSIP category and diagnosis categories whose frequency was less than 10% of the total number of first choice diagnoses were combined into an 'others' diagnosis category. The final diagnosis categories were then as follows: CTD related ILD (n=146/[490, 27.8%](#)), IPF (n=88/[490, 18.0%](#)), idiopathic NSIP (n=50/[490, 10.2%](#)), hypersensitivity pneumonitis (n=46/[490, 9.4%](#)) and others (n=160/[490, 32.7%](#)). Subsequent analyses focused on these four primary diagnosis categories.

Inter-multidisciplinary team and inter-observer agreement for first choice diagnosis

Inter-MDTM agreement and interobserver agreement (for clinicians, radiologists and pathologists) for first choice diagnosis is shown in Table 3. Overall inter-MDTM agreement for first choice diagnosis was moderate ($\kappa = 0.50$). Inter-MDTM agreement for a first choice diagnosis of IPF was good ($\kappa = 0.60$), for a first choice diagnosis of CTD related ILD also good ($\kappa = 0.64$), but poor for first choice diagnoses of idiopathic NSIP ($\kappa = 0.25$) and hypersensitivity pneumonitis ($\kappa = 0.24$). On subgroup analysis in patients in whom lung biopsy was not performed (n=48/[70, 68.6%](#)), overall inter-MDTM agreement for diagnosis was greater ($\kappa = 0.57$) with inter-MDTM agreement for first choice diagnoses of IPF, CTD related ILD and hypersensitivity pneumonitis also greater. Overall inter-observer agreement among clinicians for first choice

diagnosis was moderate ($\kappa = 0.45$) and fair between radiologists ($\kappa = 0.33$) and pathologists ($\kappa = 0.31$) (Table 3.).

Inter-multidisciplinary team and inter-observer agreement for diagnosis probabilities

There was fair to good inter-MDTM agreement on the estimation of diagnostic likelihood of the 4 most prevalent diagnoses as shown in Table 4. In particular, inter-MDTM agreement on the diagnostic likelihood of IPF was good ($\kappa_w = 0.71$, IQR 0.64-0.77), and good for CTD related ILD ($\kappa_w = 0.73$, IQR 0.68-0.78) but moderate for idiopathic NSIP ($\kappa_w = 0.41$ IQR 0.37-0.49) and fair for hypersensitivity pneumonitis ($\kappa_w = 0.29$ IQR 0.24-0.40). Subgroup analysis of inter-MDTM agreement on the estimation of diagnostic likelihood of IPF in patients without lung biopsy was good ($\kappa_w = 0.78$ IQR 0.74-0.83).

Agreement between clinicians on the probability of a diagnosis of IPF or CTD-related ILD was superior to agreement on the probability of a diagnosis of idiopathic NSIP or hypersensitivity pneumonitis. Agreement between radiologists or pathologists on the probability of a diagnosis of IPF was superior to agreement on the probability of a diagnosis of CTD-related ILD, idiopathic NSIP or hypersensitivity pneumonitis.

Subgroup analysis of inter-MDTM agreement in patients without an established diagnosis of a connective tissue disease.

At the time of patient selection, [13/70 \(18.6%\)](#) patients had an established diagnosis of a CTD (systemic sclerosis = 7, rheumatoid arthritis = 3, Sjögren's syndrome = 2, mixed connective tissue disease = 1). In order to investigate if high agreement in

these 13 cases caused a spurious increase in agreement on non-CTD diagnoses and in particular, impacting agreement on a diagnosis of IPF, a subgroup analysis was performed in the remainder of the cohort (n=57/[70, 81.4%](#)). On this analysis, although inter-MDTM agreement for a first choice diagnosis of CTD related ILD decreased ($\kappa = 0.42$), no significant change in inter-MDTM agreement was observed for a first choice diagnosis of IPF ($\kappa = 0.58$), idiopathic NSIP ($\kappa = 0.24$) or hypersensitivity pneumonitis ($\kappa = 0.23$) (Appendix, Table A2).

Diagnostic confidence for first choice diagnoses

A total of 347/490 (70.1%) first choice MDT diagnoses were made with high confidence (diagnostic likelihood 70–95% = high confidence, 100% = pathognomonic). The median prevalence of first choice MDT diagnoses made with high confidence was 67.1% (IQR 54.3-88.8). The median prevalence of first choice diagnoses made with high confidence by clinicians, radiologists or pathologists was 58.9% (IQR 52.9-71.4), 68.6% (IQR 35.7-85.7) and 72.7% (IQR 59.1-81.8) respectively (Appendix, Table A3). On subgroup analysis in the 48/[70 \(68.6%\)](#) patients who did not undergo surgical lung biopsy 237/336 (70.5%) first choice MDT diagnoses were made with high confidence. Within this subgroup the median prevalence of first choice diagnoses made with high confidence by the MDTs, clinicians or radiologists were 68.7% (IQR 52.8-87.5), 60.4% (IQR 37.5-75.0) and 66.6% (IQR 39.6-83.3) respectively.

For the diagnosis of IPF, supportive non-significant trends for higher confidence diagnoses by MDTMs (68/88, [77.3%](#)) as compared to clinicians (62/96, 64.6%) or radiologists (57/86, 66.3%) were demonstrated ($p = 0.23$). In the 22/70 ([31.4%](#)) cases that underwent surgical lung biopsy (therefore a total of 154

diagnoses rendered by 7 pathologists), 15/154 (9.7%) cases were assigned a diagnosis of IPF, of which 12/15 (80.0%) were assigned with high confidence. ____

A review of the cases where the pathologists gave a first choice diagnosis of IPF (15/154 cases, 154 = 22 cases x 7 pathologists) was performed to ascertain if, in cases where surgical lung biopsy was performed and the first choice pathologic diagnosis was IPF, the final MDTM diagnosis was usually IPF. In 6/15 cases, despite the pathologist giving a first choice diagnosis of IPF, the final MDTM first choice diagnosis was not IPF. Furthermore, in only 2/15 cases, was IPF not already suggested by either the clinician or radiologist in that MDTM (Table A5).

Formatted: Font: (Default) Arial

Evaluation of the prognostic significance of an MDTM diagnosis of IPF

On univariate Cox regression analysis, the multidisciplinary distinction between IPF and other diagnoses demonstrated non-significant trends toward greater prognostic separation (as judged by hazard ratio p values) than the clinician distinction (in 5/7 groups) or the radiologist distinction (in 4/7 groups) (Table 5). ~~On univariate Cox regression analysis, MDTM diagnosis of IPF demonstrated greater prognostic significance than clinician's diagnosis of IPF in 5/7 multidisciplinary groups, and in radiologist's diagnosis of IPF in 4/7 groups (Table 4).~~ A graphical presentation of the KM-survival curves for the categorisation of first-choice diagnosis as IPF or not IPF in at least 4/7 MDTMs, 4/7 clinicians and 4/7 radiologists is shown in Figure 1. The same analysis for pathologists' diagnosis of IPF failed to reach statistical significance for 5/7 pathologists, probably because of the small subgroup size (n=22) and low

Formatted: Font: (Default) Arial, Not Bold

Formatted: Pattern: Clear (White)

Formatted: Font: (Default) Arial, Not Bold

Formatted: Font: (Default) Arial, Not Bold

prevalence of IPF within this subgroup (15/154, pathologists' first-choice diagnoses were IPF) (Appendix, Table A4 and A5).

Formatted: Font color: Text 1, English (United States)

Discussion

We have demonstrated for the first time, that there is acceptable diagnostic agreement between multidisciplinary groups in the setting of diffuse parenchymal lung disease and that this agreement is validated by the greater prognostic significance of an IPF diagnosis made by multidisciplinary groups as compared to individual clinicians and radiologists. Furthermore, MDTMs make the diagnosis of IPF with high confidence, more frequently than clinicians or radiologists.

Since the publication of the ATS/ERS 2002 consensus statement on the classification of the ILDs, multidisciplinary evaluation of DPLD has been widely adopted as the diagnostic gold standard^{3,4}. This diagnostic approach has been investigated in a limited way in several settings. Flaherty et al. examined the formulation of diagnosis in a cohort of diffuse lung diseases against interobserver

agreement and diagnostic confidence within a single multidisciplinary team (consisting of 3 clinicians, 2 radiologists and 2 pathologists) and demonstrated that diagnostic agreement between observers improved with successive integration of clinical, radiologic and pathologic data⁵. In a second study, Flaherty et al. expanded these findings by demonstrating higher levels of agreement between academic physicians, radiologists and pathologists for diagnosis in diffuse lung disease when compared with their community counterparts⁷. Several years later, Thomeer et al. demonstrated in a cohort of patients included in an IPF trial, high accuracy for a clinical diagnosis of IPF made by 6 respiratory physicians from different European centres⁶. A limitation of these studies is that not all evaluated agreement between different MDTs for diagnosis⁵, one focuses specifically on the diagnosis of IPF⁶ and all of them predate the most recent 2013 ATS/ERS IIP update^{4-7,9}. The current study is the first to evaluate inter-MDTM agreement for diagnosis in diffuse lung disease since publication of the 2013 ATS/ERS update on the classification of the IIPs.

With the recent licensing of two new drugs for treatment of IPF in mild to moderate disease, early and accurate diagnosis of IPF is crucial. Our study demonstrates that inter-MDTM agreement for the diagnosis of IPF is good, with clinicians demonstrating only marginally inferior levels of agreement, as compared to MDTMs for this diagnosis. On subgroup analysis in patients without surgical lung biopsy, inter-MDTM agreement and interobserver agreement between clinicians for the diagnosis of IPF improved and were again, almost equal. As evidence-based diagnostic criteria for IPF are now clearly defined and in many IPF patients, are relatively easily applied, this near parity of agreement between MDTMs and between clinicians is not surprising^{4,10}. For example, in a patient with classic appearances for usual interstitial pneumonia (UIP) on HRCT, a rapidly progressive disease course and

no identifiable triggers, multidisciplinary discussion is unlikely to alter a clinician's initial impression of IPF. Our findings suggest that formal multidisciplinary input may not be necessary in every case of suspected IPF if expert clinical evaluation is available, which may be of particular relevance to centres with limited access to appropriate radiology or pathology expertise¹³.

A strength of our findings is that the greater agreement on multidisciplinary diagnosis is mirrored by the trend toward greater prognostic ~~significance-separation~~ of a multidisciplinary ~~diagnosis-distinction between~~ IPF and other diagnoses when compared to the clinicians' or radiologists' distinction. To demonstrate this, patients were necessarily selected from 2010 to allow an analysis of 5-year survival. An added advantage of this approach was that these patients were referred in a pre-antifibrotic drug era, therefore mortality differences between IPF and non-IPF patients were not confounded by antifibrotic therapy. The fact that trends are present but inconclusive may reflect powering limitations but may also indicate that multidisciplinary discussion adds prognostic value in the sub-group of patients in which there is significant diagnostic uncertainty

Formatted: Font: (Default) Arial, Not Bold

In contrast to those with IPF, many patients with non-IPF interstitial lung diseases are not covered by evidence-based diagnostic guidelines therefore diagnosis is driven by clinical reasoning and analysis of all available data in a multidisciplinary setting. In these situations, a level of disagreement between MDTMs is predictable and borne out by the poor level of inter-MDTM agreement in the current study for diagnoses of NSIP and hypersensitivity pneumonitis^{4,14-16}. In particular, it is well recognised that the diagnosis of hypersensitivity pneumonitis can be challenging as it relies on an array of data, none of which is definitive when considered in isolation¹⁵, and at least on HRCT, patterns of NSIP, UIP or organising pneumonia

may be the sole expressions of this disease¹⁶⁻¹⁹. The poor MDTM agreement for the diagnosis of hypersensitivity pneumonitis in the current study highlights an urgent need for international consensus on what hypersensitivity pneumonitis actually is.

Our use of the weighted kappa to investigate inter-MDTM and interobserver agreement on diagnostic probabilities is similar to other studies of diagnostic agreement, but warrants particular consideration⁹. Converting the diagnostic likelihoods to a 5-point probability scale enabled examination of the range of diagnostic likelihoods from minimal likelihood to pathognomonic using the weighted kappa. As excluding IPF is as important as making a diagnosis of IPF, this methodology has allowed us to demonstrate that there is good agreement on the likelihood of IPF and as stated previously, this reflects consistent application of established diagnostic guidelines for this disease. In contrast, MDTM agreement on the likelihood of hypersensitivity pneumonitis was poor, reinforcing the view that MDTMs were unclear on how this diagnosis is made.

Two separate observations from our study warrant further discussion. First, 13 patients had an established diagnosis of a connective tissue disease at presentation. However, following MDT evaluation, a CTD diagnosis was constructed in an additional 8 cases or more by five of the seven MDTMs based on presenting clinical symptoms and serology. Separating patients with IIP from those with CTD related ILD can be challenging - some patients present with subtle clinical features or serological abnormalities which suggest an autoimmune process but do not meet established criteria for a specific CTD²⁰⁻²². Recently an ERS/ATS task force was formed in order to establish consensus on how to classify these patients and a set of diagnostic criteria have been suggested²³⁻²⁵. Following removal of patients with an established diagnosis of CTD, our subgroup analysis demonstrated acceptable levels of

agreement on CTD related ILD (in contrast to hypersensitivity pneumonitis), underlying the importance of reaching international consensus on definitions. Furthermore, the high frequency of CTD related ILD diagnoses made in the current study highlights the importance of formal rheumatology input within the MDTM, which might necessarily include face-to-face rheumatologic consultation with patients, suspected of having an undiagnosed connective tissue disease.

Second, it has previously been suggested that dynamic exchanges of clinical, radiologic and pathologic information between experts in a multidisciplinary process results in higher confidence diagnoses⁵. In the current study however, the proportion of high confidence diagnoses (70% or greater) assigned following MDTM evaluation did not increase when compared to the proportion of high confidence diagnoses being assigned by the individual components of the MDTM. In the majority of MDTMs, high confidence diagnoses were more frequently assigned by radiologists and pathologists, when compared to MDTM diagnoses or diagnoses assigned by clinicians. As radiologists and pathologists in the current study did not have access to clinical information, their interpretation was based almost entirely on pattern recognition, which might conceivably result in more confident, but not necessarily more accurate diagnoses. Although one benefit of the multidisciplinary process is that diagnoses may be challenged and must be publically defended, it is possible that extra discussion creates more difficulty in some cases that initially seem straightforward when evaluated by individuals in isolation. It must be highlighted however, that in the specific case of IPF, MDTMs made the diagnosis of IPF with high confidence more frequently than clinicians or radiologists.

Our methodology has some limitations. As patients were selected from a pre-antifibrotic drug era, IPF was possibly not as prevalent in our study group as it would

be now at most referral centres which expend more time in evaluating patients for approval for anti-fibrotic treatment. However as discussed, this allowed us to evaluate the veracity of MDTM diagnosis for IPF, against outcome. Second, unlike a real-world multidisciplinary process, none of the observers had face-to-face consultation with patients and therefore did not have the benefit of obtaining a clinical history or performing physical examination first-hand. It is difficult to characterize patients, a clinician's impression might conceivably be influenced by direct contact with the patient in ways, which are not easily and objectively quantified. Equally important, but also difficult to assess, is the impact of individual personalities on MDTM diagnosis. Arguably a consensus diagnosis at multidisciplinary evaluation might sometimes reflect the strongest voice in the room rather than represent true agreement between participants. Lastly, the wide range of proportions of high confidence diagnoses demonstrated between the different MDTM's seen in our study may reflect cultural influences on the dynamics of multidisciplinary evaluation. These less tangible influences could be the focus of further studies.

In conclusion, our study demonstrates that diagnostic agreement between MDTMs is superior to interobserver agreement between clinicians, radiologists and pathologists in the setting of DPLD. In particular, inter-MDTM agreement for a diagnosis of IPF is good and validated by the greater prognostic significance of an IPF diagnosis made by multidisciplinary groups as compared to individual specialists. Poor levels of inter-MDTM agreement were demonstrated for NSIP and hypersensitivity pneumonitis, the latter of which may relate to the lack of evidence-based guidelines for diagnosing this disease.

Research in context

Evidence before this study

We performed a PubMed search on 17th January 2014 using the search terms “diffuse parenchymal lung disease”, “idiopathic pulmonary fibrosis”, “idiopathic interstitial pneumonias”, “interobserver agreement”, “diagnosis” and “multidisciplinary team” for the period between January 2000 and January 2014. This search was extended to 1st December 2015 during the writing of the manuscript. Our search was restricted to publications written in English. We identified 7 key publications which were pertinent to our study^{3-7,9,10}. Of these, 4 studies of observer agreement in setting of diffuse parenchymal lung disease were identified^{5-7,9}. All of these studies however predated the most up to date ATS/ERS update on the classification of the idiopathic interstitial pneumonias⁴, and not all of them evaluated diagnostic agreement between multidisciplinary teams but rather focused on diagnostic agreement between individual observers.

Added value of this study

This study is the first evaluation of inter-multidisciplinary team agreement for diagnosis in the setting of diffuse parenchymal lung disease since the updated 2013 ATS/ERS classification of the idiopathic interstitial pneumonias and the 2011 ATS/ERS/JRS/ALAT guidelines for the diagnosis and management of idiopathic pulmonary fibrosis.

Implications of all the available evidence

Our study demonstrates that diagnostic agreement between MDTMs is superior to interobserver agreement between clinicians, radiologists and pathologists in the setting of DPLD. Specifically in IPF, MDTMs have a higher level of agreement on

diagnosis, assign diagnosis with higher confidence more frequently, and provide diagnoses that have greater prognostic significance than clinicians or radiologists in the majority of groups.

This is of particular importance, as accurate and consistent diagnosis of idiopathic pulmonary fibrosis is needed if clinical outcome is to be optimised. In contrast, inter-multidisciplinary team agreement for a diagnosis of hypersensitivity pneumonitis is poor, highlighting an urgent need for standardised diagnostic guidelines for this disease entity.

Conflicts of interest:

AUW received personal fees from Intermune, Roche, Bayer and Gilead. DV received personal fees from Roche, Intermune and Boehringer Ingelheim. HN is an investigator for clinical trials by Intermune, Roche, Boehringer-Ingelheim, Sanofi, and Centocor. HT received personal fees from Abbott Japan Co. Ltd, Actelion Pharmaceuticals Japan Ltd, Ashai Kasei Pharmaceutical Corporation, Astellas Pharmaceutical Incorporated, Astra Zeneca, Bayer, Boehringer Ingelheim, and Chugai Pharmaceuticals. JF received personal fees from Astellas Pharmaceutical Incorporated, Pathology Institute Corporation, Chugai Pharmaceuticals and Sakura Finetek Japan. AGN received personal fees from Sanofi, Intermune, Boehringer Ingelheim and Actelion. KRF received personal fees from Boehringer Ingelheim, Genentech, Ikaria, Immuneworks, Veracyte, Roche, Gilead, Biogen, Afferent, Aeolus, Pharmakea and grants from Boehringer Ingelheim, Genentech, Roche and Afferent. DMH received personal fees from Boehringer Ingelheim, Sanofi, AstraZeneca, Roche and Glaxo Smith Kline. The other authors declared no conflicts of interest.

Author contributions

LFW, AUW, DMH, SD - Study concept, data collection, analysis and manuscript writing. JJ, MK, CS - Data collection. AGN, LE – Data collection and evaluation of digitized pathology data. KRF, JLM – Manuscript review and editing. Scoring of MDT data. All remaining authors made up the 7 multidisciplinary teams from France, Italy, Netherlands, Japan, Portugal, United Kingdom, Denmark

1. King TE, Jr., Bradford WZ, Castro-Bernardini S, et al. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 2014; **370**(22): 2083-92.
2. Richeldi L, du Bois RM, Raghu G, et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med* 2014; **370**(22): 2071-82.
3. American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. This joint statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was adopted by the ATS board of directors, June 2001 and by the ERS Executive Committee, June 2001. *Am J Respir Crit Care Med* 2002; **165**(2): 277-304.
4. Travis WD, Costabel U, Hansell DM, et al. An official American Thoracic Society/European Respiratory Society statement: Update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. *Am J Respir Crit Care Med* 2013; **188**(6): 733-48.

5. Flaherty KR, King TE, Jr., Raghu G, et al. Idiopathic interstitial pneumonia: what is the effect of a multidisciplinary approach to diagnosis? *Am J Respir Crit Care Med* 2004; **170**(8): 904-10.
6. Thomeer M, Demedts M, Behr J, et al. Multidisciplinary interobserver agreement in the diagnosis of idiopathic pulmonary fibrosis. *Eur Respir J* 2008; **31**(3): 585-91.
7. Flaherty KR, Andrei AC, King TE, Jr., et al. Idiopathic interstitial pneumonia: do community and academic physicians agree on diagnosis? *Am J Respir Crit Care Med* 2007; **175**(10): 1054-60.
8. Tomassetti S, Piciucchi S, Tantalocco P, Dubini A, Poletti V. The multidisciplinary approach in the diagnosis of idiopathic pulmonary fibrosis: a patient case-based review. *Eur Respir Rev* 2015; **24**(135): 69-77.
9. Aziz ZA, Wells AU, Hansell DM, et al. HRCT diagnosis of diffuse parenchymal lung disease: inter-observer variation. *Thorax* 2004; **59**(6): 506-11.
10. Raghu G, Collard HR, Egan JJ, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011; **183**(6): 788-824.
11. Walsh SL, Calandriello L, Sverzellati N, Wells AU, Hansell DM, Consort UIPO. Interobserver agreement for the ATS/ERS/JRS/ALAT criteria for a UIP pattern on CT. *Thorax* 2015.
12. Investigators P. Value of the ventilation/perfusion scan in acute pulmonary embolism. Results of the prospective investigation of pulmonary embolism diagnosis (PIOPED). *JAMA* 1990; **263**(20): 2753-9.
13. Jo HE, Corte TJ, Moodley Y, et al. Evaluating the interstitial lung disease multidisciplinary meeting: a survey of expert centres. *BMC Pulm Med* 2016; **16**(1): 22.

14. Travis WD, Hunninghake G, King TE, Jr., et al. Idiopathic nonspecific interstitial pneumonia: report of an American Thoracic Society project. *Am J Respir Crit Care Med* 2008; **177**(12): 1338-47.
15. Lacasse Y, Selman M, Costabel U, et al. Clinical diagnosis of hypersensitivity pneumonitis. *Am J Respir Crit Care Med* 2003; **168**(8): 952-8.
16. Silva CI, Muller NL, Lynch DA, et al. Chronic hypersensitivity pneumonitis: differentiation from idiopathic pulmonary fibrosis and nonspecific interstitial pneumonia by using thin-section CT. *Radiology* 2008; **246**(1): 288-97.
17. Lacasse Y, Girard M, Cormier Y. Recent advances in hypersensitivity pneumonitis. *Chest* 2012; **142**(1): 208-17.
18. Churg A, Sin DD, Everett D, Brown K, Cool C. Pathologic patterns and survival in chronic hypersensitivity pneumonitis. *Am J Surg Pathol* 2009; **33**(12): 1765-70.
19. Silva CI, Churg A, Muller NL. Hypersensitivity pneumonitis: spectrum of high-resolution CT and pathologic findings. *AJR Am J Roentgenol* 2007; **188**(2): 334-44.
20. Fischer A, du Bois R. Interstitial lung disease in connective tissue disorders. *Lancet* 2012; **380**(9842): 689-98.
21. Tzelepis GE, Taya SP, Moutsopoulos HM. Occult connective tissue diseases mimicking idiopathic interstitial pneumonias. *Eur Respir J* 2008; **31**(1): 11-20.
22. Mittoo S, Gelber AC, Christopher-Stine L, Horton MR, Lechtzin N, Danoff SK. Ascertainment of collagen vascular disease in patients presenting with interstitial lung disease. *Respir Med* 2009; **103**(8): 1152-8.
23. Fischer A, West SG, Swigris JJ, Brown KK, du Bois RM. Connective tissue disease-associated interstitial lung disease: a call for clarification. *Chest* 2010; **138**(2): 251-6.

24. Vij R, Noth I, Strek ME. Autoimmune-featured interstitial lung disease: a distinct entity. *Chest* 2011; **140**(5): 1292-9.
25. Kinder BW, Collard HR, Koth L, et al. Idiopathic nonspecific interstitial pneumonia: lung manifestation of undifferentiated connective tissue disease? *Am J Respir Crit Care Med* 2007; **176**(7): 691-7.

|

Variable	Value
Age (Years)	60.9 ± 15.5
Gender (M : F)	24:46
Smoking (Never/ever/current)	38:24:8
Established CTD history	13
Biopsied (Y/N)	22/48
DL _{co} (% Predicted)	44.8 ± 14.5
FEV ₁ (% Predicted)	73.0 ± 20.5
FVC (% Predicted)	79.0 ± 19.6

Table 1. Patient demographics. CTD = connective tissue disease. Where appropriate, values are expressed as means ± SD.

|

Diagnosis	Number (n=490)
Connective tissue disease related ILD	146 (29.8%)
IPF	88 (18.0%)
Hypersensitivity pneumonitis	46 (9.4%)
NSIP	42 (8.6%)
Unclassifiable ILD	38 (7.8%)
Other	37 (7.6%)
Sarcoidosis	20 (4.1%)
Drug-related ILD	18 (3.7%)
Smoking-related ILD	16 (3.3%)
Occupational lung disease	14 (2.9%)
Pleuroparenchymal fibroelastosis	10 (2.0%)
OP/NSIP overlapping disease	8 (1.6%)
Organizing pneumonia	7 (1.4%)

Table 2. First choice diagnoses given by 7 multidisciplinary teams for 70 cases of diffuse lung disease. ILD = interstitial lung disease, NSIP =non-specific interstitial pneumonia, OP = organising pneumonia.

	MDTM (κ)	Clinicians (κ)	Radiologists (κ)	Pathologists (κ)
Overall Total (No biopsy)	0.50 (0.57)	0.45 (0.50)	0.33 (0.31)	0.31
IPF Total (No biopsy)	0.60 (0.70)	0.59 (0.71)	0.46 (0.42)	0.46
NSIP Total (No biopsy)	0.25 (0.25)	0.19 (0.19)	0.25 (0.25)	0.23
CTD Total (No biopsy)	0.64 (0.73)	0.57 (0.62)	0.10 (0.11)	0.22
HP Total (No biopsy)	0.24 (0.31)	0.25 (0.38)	0.27 (0.22)	0.20

Table 3. Unweighted Kappa values (κ) for inter-MDTM agreement for overall diagnoses and for individual diagnoses of idiopathic pulmonary fibrosis (IPF), connective tissue disease related interstitial lung disease (CTD), non-specific interstitial pneumonia (NSIP) and hypersensitivity pneumonitis (HP). Unweighted Kappa values for inter-observer agreement among clinicians, radiologists, and pathologists are also shown. Figures in parentheses are Kappa values in non-

biopsied cases, n=48 (see text). *22/70 (31.4%) cases had surgical lung biopsy performed.

	MDTM (κw)	Clinicians (κw)	Radiologists (κw)	Pathologists (κw)
IPF	0.71 (0.64-0.77)	0.72 (0.67-0.76)	0.60 (0.46-0.66)	0.58 (0.45-0.66)
CTD	0.73 (0.68-0.78)	0.76 (0.70-0.78)	0.17 (0.08-0.31)	0.21 (0.06-0.36)
NSIP	0.42 (0.37-0.49)	0.31 (0.27-0.41)	0.32 (0.26-0.41)	0.30 (0.00-0.53)
HP	0.29 (0.24-0.40)	0.42 (0.30-0.47)	0.35 (0.29-0.43)	0.26 (0.10-0.45)

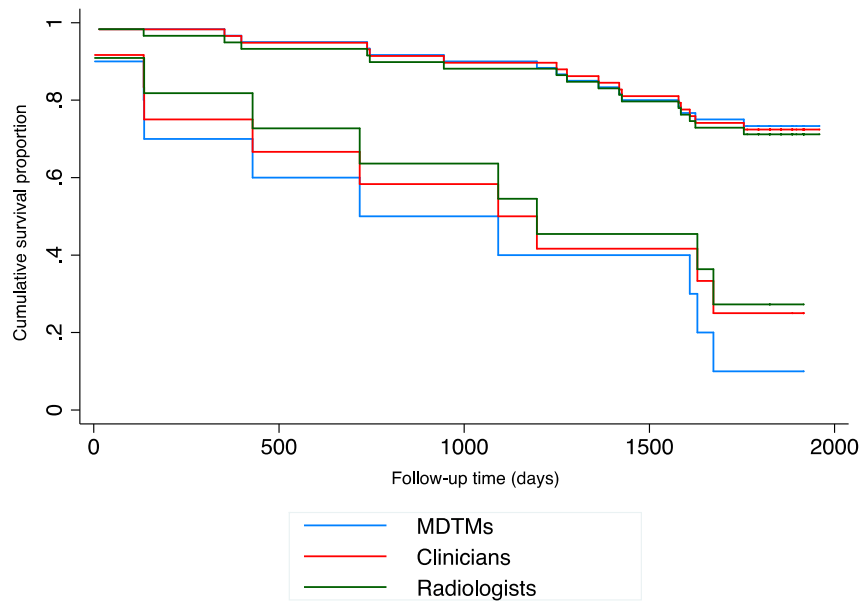
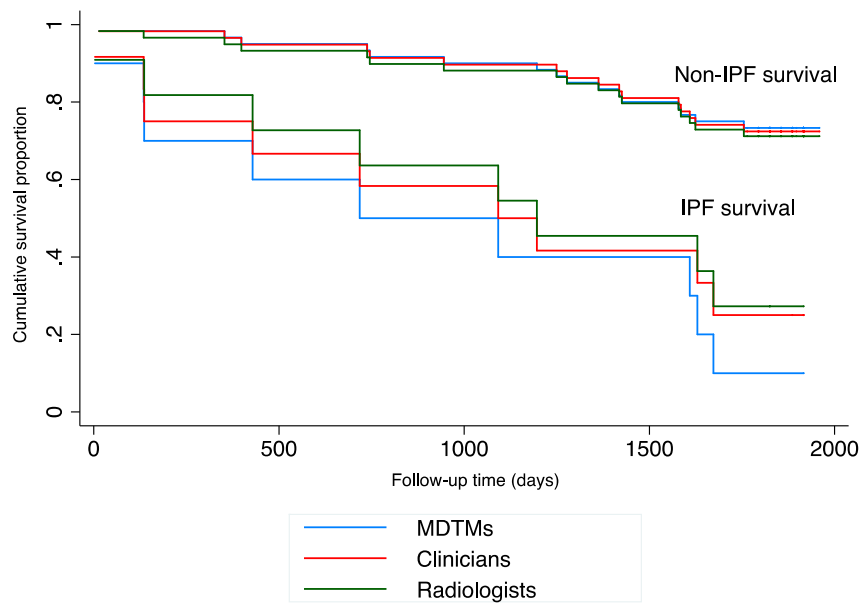
Table 4. Weighted Kappa values (**κw**) for estimation of diagnostic likelihood for individual diagnoses of idiopathic pulmonary fibrosis (IPF), connective tissue disease related interstitial lung disease (CTD), non-specific interstitial pneumonia (NSIP) and hypersensitivity pneumonitis (HP), expressed as medians with interquartile ranges.

	MDTM (HR, p-value, 95%CI)	Clinicians (HR, p-value, 95%CI)	Radiologists (HR, p-value, 95%CI)
Team 1	2.67, p = 0.016, 1.21-6.02	2.09, p = 0.085, 0.90-4.86	2.80, p = 0.021, 1.17-6.73
Team 2	3.44, p = 0.003, 1.54-7.68	2.95, p = 0.008, 1.33-6.59	4.08, p = 0.001, 1.84-9.04
Team 3	5.30, p < 0.001, 2.26-12.41	3.75, p = 0.002, 1.65-8.51	2.78, p = 0.030, 1.11-6.97
Team 4	3.99, p = 0.006, 1.49-10.66	3.34, p = 0.007, 1.38-8.00	4.49, p = 0.003, 1.71-12.29
Team 5	2.61, p = 0.025, 1.12-6.06	2.03, p = 0.100, 0.87-4.69	2.58, p = 0.033, 1.08-6.21
Team 6	3.36, p = 0.007, 1.40-8.07	4.14, p = 0.002, 1.72-9.97	2.11, p = 0.082, 0.91-4.89
Team 7	2.43, p = 0.030, 1.09-5.41	2.96, p = 0.007, 1.43-6.55	1.28, p = 0.583, 0.53-3.06

Formatted Table

Formatted: Left, None, Space
Before: 0 pt, Don't keep with next,
Don't keep lines together

Table 5. Univariate Cox regression analysis for mortality according to MDTM, clinician and radiologists diagnoses of IPF versus not IPF. (MDTM = multidisciplinary team meeting). The results for the MDTMs, clinicians and radiologists are based upon the full patient cohort (n=70). *The results for the pathologists are for the subgroup of 22/70 (31.4%) patients who underwent surgical lung biopsy

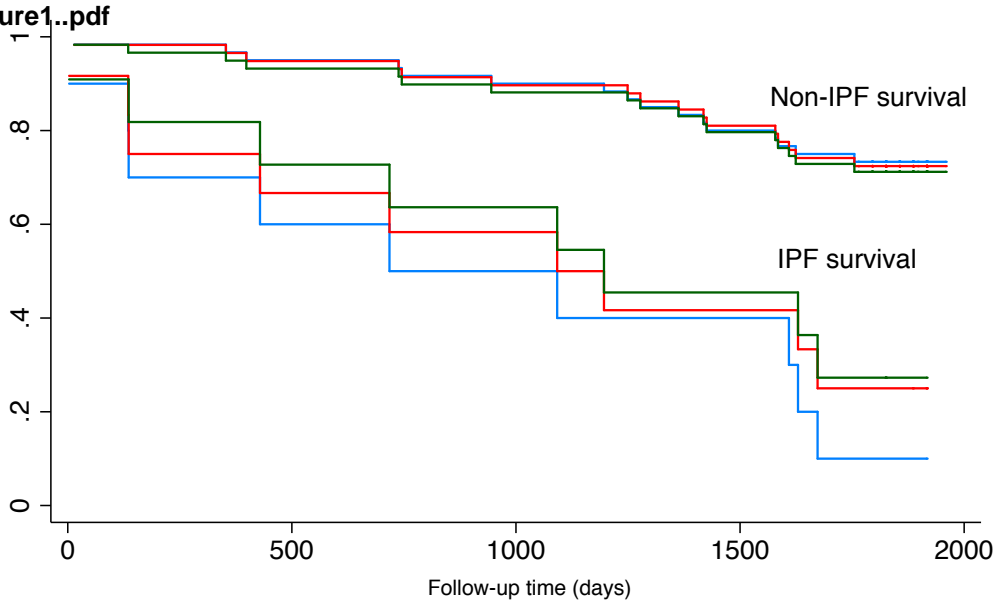


Formatted: Font: (Default) Arial

Figure 1. Kaplan-Meier survival curve demonstrating survival differences between patients assigned a diagnosis of IPF or not IPF. The entire cohort was separated into cases where at least 4/7 MDTMs assigned a first-choice diagnosis of IPF (Blue KM curve, HR 6.26, $p < 0.0001$, 95% CI 2.72-14.33), at least 4/7 clinicians assigned a first-choice diagnosis of IPF (red KM curve, HR 4.43, $p < 0.0001$, 95% CI 1.94-10.01), and at least 4/7 radiologists assigned a first-choice diagnosis of IPF (green KM curve, HR 3.76, $p = 0.002$, 95% CI 1.61-8.75). MDTM = multidisciplinary team meeting. See Appendix, Figure 7A for number at risk table.

figure1..pdf

Cumulative survival proportion



MDTMs
Clinicians
Radiologists

appendix_r4.pdf

[Click here to download Necessary Additional Data: APPENDIX_R4.pdf](#)

appendixfigure1.pdf

[Click here to download Necessary Additional Data: AppendixFigure1.pdf](#)

King's College Hospital

NHS Foundation Trust

King's College Hospital NHS Foundation Trust

King's College Hospital
Denmark Hill
London SE5 9RS

Direct tel: 020 3299 3526
Direct fax: 020 3299 3157
Email: slfwalsh@gmail.com

17th March 2016,
Diana Stanley,
Editorial Office
Lancet Respiratory Medicine

Re : Manuscript THELANCETRM-D-16-0035

Dear Professor Stanley,

We are delighted to be given the opportunity to revise our manuscript. We hope that we have addressed all of the your comments to your satisfaction and in particular thank you for taking into consideration our concerns regarding how the data pertaining to the pathologists in our study should be displayed.

Specific Editorial points to be addressed:

1. We understand your concerns with adding the pathologist survival data to the main paper and table. We wonder though if it would be worth doing the analysis in the appendix so readers can see if they want to?

Many thanks to taking our concerns regarding this into consideration. We have added this data to the Appendix (Table A5) as suggested.

2. You say that a total of 22/70 (34.1%) of cases underwent surgical lung biopsy. Can you clarify why the others weren't biopsied - confident diagnosis or not suitable candidates?

Based upon a review of the patients electronic patient records, in all 48/70 cases who did not undergo surgical lung biopsy, a confident diagnosis was made without the need to proceed to lung biopsy. We have added this to the manuscript (Results, patient population)

3. Please clarify that clearer you weren't using the diagnostic classifications of PE from the PIOPED study but the criteria for diagnostic likelihood. I think this is where confusion lies.

Thank you. We have modified the manuscript accordingly.

4. Please add to appendix and mention in main paper:
we reviewed the cases where the pathologists gave a first choice diagnosis of IPF (15/154 cases, 154 = 22 cases x 7 pathologists). We have now included this in tabular form in the appendix (Table A5). You will see that in only 2 of these 15 cases was IPF not already given as a first choice diagnosis by either the clinician or radiologist (marked with *). In 6/15 cases, despite the pathologist giving a first choice diagnosis of IPF, the final MDTM first choice diagnosis was not IPF.

Many thanks for this suggestion. We have added (Results, diagnostic confidence for first choice diagnoses):

“A review of the cases where the pathologists gave a first choice diagnosis of IPF (15/154 cases, 154 = 22 cases x 7 pathologists) was performed to ascertain if, in cases where surgical lung biopsy was performed and the first choice pathologic diagnosis was IPF, the final MDTM diagnosis was usually IPF. In 6/15 cases, despite the pathologist giving a first choice diagnosis of IPF, the final MDTM first choice diagnosis was not IPF. Furthermore, in only 2/15 cases, was IPF not already

suggested by either the clinician or radiologist in that MDTM (Table A5)."

5. Please add some information in a table or in appendix about age, gender, smoking history, history of established connective tissue disease and number of cases that underwent lung biopsy and if you have it FEV₁ for the patient population so the reader can get an idea of what it was like.

We have included a table (Table A2) in the appendix which shows this data, as well as the % predicted DL_{co}, FEV₁ and FVC.

General editorial comments to be addressed:

1. Please check with your co-authors, and confirm, that all names are spelt correctly, and affiliations listed correctly. We cannot guarantee that we will be able to correct names and affiliations after publication of your article.

Many thanks – we have checked this.

2. The study title should have a descriptor—ie, randomised trial, case-control study, prospective analysis, population-based study

Many thanks – we have adjusted the title as requested (Multicentre evaluation of multidisciplinary team meeting agreement on diagnosis in diffuse parenchymal lung disease: a case cohort study)

3. Please supply (after author names on the title page) one preferred degree per author and indicate in the authorship if any authors are full professors.

Many thanks – we have added a title page as instructed

4. Please confirm that your study conforms to the relevant guidelines by completing and returning the checklist.

STROBE - Observational studies —

[http://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(07\)61602-X/fulltext](http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(07)61602-X/fulltext)

Many thanks – we have included the STROBE checklist with the statement “We have followed this checklist and we can confirm that our manuscript stratifies the recommendations in each section detailed below”.

5. It is Lancet style is to give actual numbers (nominator and denominator) together with percentages.

Many thanks – we have made these changes.

6. Lancet style is to provide exact p values, unless $p < 0.0001$ (if this is the case, then please revise to the latter). You need only give the p-value to 2 decimal places for non-significant results (ie $p = 0.87$).

Many thanks – we have made these changes

7. The Lancet group are very supportive of protocol-based research and so have recently decided to encourage authors to post the protocol document on a publicly accessible website; a margin link to the website will then be put in the paper. Would you like to do this for your protocol? If so, please send us the protocol link with your final corrections. Please note that if you do wish to do this then the weblink should not be temporary.

Not applicable

8. Lancet style is to have a 'Role of the funding source' at the end of the methods. The following points need to be addressed in the "Role of the funding source" statement:

- * The role of the sponsors in the study design.
- * The role of the sponsors in the collection, analysis, or interpretation of the data.
- * The role of the sponsors in the writing of the report.
- * Those who had access to the raw data (by author initials).

If the funding source had no role then this should be stated. Please also add to this section (if true): "The corresponding author had full access to all of the data and the final responsibility to submit for publication."

Many thanks – we have added the following after the methods section:

“Role of the funding source:

The sponsors of the study did not have any role in the design, data collection, analysis and interpretation, nor in the writing of the report. The corresponding

author had full access to all of the data and the final responsibility to submit for publication."

9. If you have claimed a first, please reword to: "To our knowledge... this is the first time...", since you can never be 100% sure.

Not applicable

10. We require TWO author signatures from every authors listed. All authors need to sign this form: <http://download.thelancet.com/flatcontentassets/authors/tlrm-author-signatures.pdf> AND each author must additionally complete and return this form: <http://download.thelancet.com/flatcontentassets/authors/icmje-coi-form.pdf>. You can return these by email if it is easier.

We currently have all the author forms except for Dr Poletti. We don't have any conflict of interest forms.

We have attached these documents.

11. Please add a conflict of interest statement to the end of your paper, as per Lancet style. These statements should exactly match those given on your signed conflict of interest forms. If there are none then please state "The authors declared no conflicts of interest" or "The other authors declared no conflicts of interest."

We have added this to the end of the manuscript.

12. Please add an Author contributions section to the end of your paper before the references, as per Lancet style. These statements should exactly match those given on your signed forms. Authors should be referred to by their initials in this section.

We have added this after the conflict of interest statement and before the references section.

13. References should be in Vancouver style. For references with six authors or fewer, all authors should be listed. For those with seven or more authors, then just the first three authors and 'et al' should be listed. Please ensure that you do not use endnotes and footnotes in Word as they are difficult to edit.

Completed

14. All web references should have the date they were last accessed.

Not applicable

15. All abstracts and references in press should be updated with DOIs or page numbers as appropriate. For papers listed in references that are "in press" we need to see a galley proof and/or letter from the publisher stating that it is "in press" as well as (where known) the full expected citation (ie, publication date/volume/issue etc). References that are "submitted" should be removed and citations in the text replaced with "(unpublished data; authors)".

Completed

16. Please supply all tables in an editable format as Word files (not excel or fdf/pdf). Each row of data should be in a separate line. Please ensure that rows and columns are not tabbed; data should be entered in cell form.

Completed

17. Please supply figures as editable high-resolution EPS format, exported directly from your statistical package if possible, rather than embedded in a Word file. (TIF for photographs).

Not applicable

18. Please add number at risk in each group for each time point on your K-M curves.

Completed.

19. It is not Lancet policy to edit or style supplementary material for the web; however, this material will still be hosted on our website as a pdf of the author supplied file. Please style your supplementary material as per the guidelines found at <http://www.thelancet.com/lancet-information-for-authors/web-extra-guidelines>. Please note that we will be unable to correct any errors in the webappendix following publication; as such, please check carefully when submitting. Please supply the webappendix as a single PDF file, with the pages paginated.

We have submitted the appendix as a PDF file.

20. Please ensure that main outcome measures include a result for each group plus a point estimate (eg, RR, HR) with a measure of precision (95% CI) for the difference between groups.

Completed.

Reviewers' comments:

Reviewer #1: The authors have addressed this reviewer's comments and amended the manuscript accordingly.

Many thanks for your work.

Reviewer #4:

the curves for Figure 1 might still confuse a reader, and if the upper ones could be labeled as non-IPF survival and the lower ones as IPF survival curves (survival differences for the two cohorts per the 3 different diagnostic determination methods), that would make it a bit clearer for readers.

Many thanks for this suggestion. We have added these labels to the graph as suggested.

Statistical review comments:

We would like to thank the new statistical reviewer for reviewing our manuscript so promptly under difficult circumstances.

Major comment 1: the definition of follow up time can be further clarified, for example, by adding “minimum of” before the phrase “date of death, date the patient was last known to be alive or 1st June 2015”.

Many thanks for this comment. We modified the manuscript accordingly.

Major comment 2: no additional comment.

Major comment 3: the interpretation of the reported hazard ratios should be added. For example, hazard ratio (IPF vs. not IPF). I don't think the results on checking the PH assumptions were shown, so need to at least add a phrase "(data not shown)" in the description.

Many thanks for this comment. We have added these suggestions to the manuscript (Methods section, statistical analysis paragraph 4).

Major comment 4: no additional comment.

Major comment 5: I think the requested information was provided in Table 2A, so this just needs to be pointed out in the response.

The suggested detail has been reported in Table 2A. This also includes the variables suggested by the editors in the editorial comments (comment 5)

Major comments 6 & 7: no additional comment.

Major comment 8: suggest changing the x-axis label to be something more specific. For example "Time since referral (days)". The legend seems to be incomplete. What lines are for the IPF and what lines are for not IPF?

Many thanks for this suggestion. We have modified the figure accordingly and provided labels for the graph lines (IPF, for MDTMs, clinicians and radiologists, and Not-IPF, for MDTMs, clinicians and radiologists).

Further comments:

I find no statistical evidence presented to support the following statement "On univariate Cox regression analysis, MDTM diagnosis of IPF demonstrated greater prognostic significance than clinician's diagnosis of IPF in 5/7 multidisciplinary groups, and in radiologist's diagnosis of IPF in 4/7 groups". What basis did the author

use to conclude a greater prognostic significance between MDTM vs. clinicians or radiologists' diagnosis? The p-value of testing $HR=0$? This needs to be at least clarified if not corrected.

Many thanks for this comment

The purpose for examining the prognostic separation resulting from MDTM, clinicians' and radiologists' diagnoses of IPF or not IPF was to determine whether major differences in prognostic separation exist. Had this been the case, routine multidisciplinary discussion in all cases would be justified. The fact that trends are present but inconclusive may reflect powering limitations but may also indicate that multidisciplinary discussion adds prognostic value in the sub-group of patients in which there is significant diagnostic uncertainty.

We agree that this statement needs clarification and actually a correction. The term prognostic "significance" is confusing when what we mean is prognostic "separation". The text is modified as follows (Results, Evaluation of the prognostic significance of an MDTM diagnosis of IPF)

"On univariate Cox regression analysis, the multidisciplinary distinction between IPF and other diagnoses demonstrated greater prognostic separation (as judged by hazard ratio p values) than the clinician distinction (in 5/7 groups) or the radiologist distinction (in 4/7 groups)."

We have modified the discussion (paragraph 4) as follows:

"A strength of our findings is that the greater agreement on multidisciplinary diagnosis is mirrored by the trend toward greater prognostic separation of a multidisciplinary distinction between IPF and other diagnoses when compared to the clinicians' or radiologists' distinction. To demonstrate this, patients were necessarily selected from 2010 to allow an analysis of 5-year survival. An added advantage of this approach was that these patients were referred in a pre-

antifibrotic drug era, therefore mortality differences between IPF and non-IPF patients were not confounded by antifibrotic therapy. The fact that trends are present but inconclusive may reflect powering limitations but may also indicate that multidisciplinary discussion adds prognostic value in the sub-group of patients in which there is significant diagnostic uncertainty”

We have also modified the abstract by replacing the word significance with “separation”.

The Cox regression analysis presented in Figure 1 is questionable. If I understand it correctly, the total diagnoses presented in the analysis were not for independent cases, but rather for 70 independent cases repeatedly assessed by multiple MDTMs or multiple individual experts. The inferences on the HRs by MDTMs in Table 4 seem to be appropriate because each MDTM assessed 70 independent cases.

The purpose of this analysis was simply to provide a graphical demonstration of the greater prognostic separation of the cohort based upon MDTMs’ diagnosis (of IPF or not IPF) versus clinicians’ diagnosis or radiologists’ diagnosis (of IPF versus not IPF). I had originally graphed all 7 MDTMs, clinician’s and radiologist’s but the graph was very cluttered and it was difficult to visually inspect. Therefore as a compromise, I separated the cohort into IPF versus not IPF based upon what MOST of the MDTMs, clinicians or radiologists diagnosed (4/7). I stress however that this figure was for illustrative purposes only and does not impact the message demonstrated by Table 4. If the reviewer has a strong view on whether this figure should/should not be included we would be very happy to consider that view.

Simon LF Walsh MD MRCP FFRRCSI

Consultant Thoracic Radiologist

King’s College Hospital

Athol U Wells MD FRCP FRCR

Professor of Respiratory Medicine
Royal Brompton Hospital

David M Hansell MD FRCP FRCR

Professor of Thoracic Imaging
Royal Brompton Hospital